Online Journal of Public Health Informatics

High-quality research and innovation in the field of public health informatics
Volume 18 (2026) ISSN 1947-2579 Editor in Chief: Edward K. Mensah, PhD, MPhil

Contents

Viewpoints

Bias in AI in Epic Electronic Medical Records and Its Impact on Executive Communication: Conceptual Analysis (e85584)
Naomi Pemberton. 5

Transforming Pediatric Care Through AI: Bridging the Digital Divide in Health Informatics (e82108)
Mercy Mbogori-Kairichi, R Snead, Julia Marczika, Radha Ambalavanan, Gideon Towett, Alex Malioukis. 10

Reviews

Trend in Devices and Digital Tools for Remote Consultation From Medical Providers to Specialists: Scoping Review (e87559)
Risa Hara, Yosuke Hirakawa, Yuka Sugawara, Masao Iwagami, Ryota Inokuchi, Masaomi Nangaku. 19

Adaptation of Social Health Care Tools From Paper-Based to Web-Based Formats: Scoping Review (e88631)
Charlèss Dupont, Eva Savolainen, Petter Fjällström, Malin Eneslätt. 35

Safety-Oriented Evaluation of Large Language Models in Health Care: Guideline-Informed Systematic Review (e97240)
Hikaru Matsuoka, Takayuki Takahashi, Takayuki Semitsu, Keiko Matsuoka, Ryoma Hayashi. 49

Evaluating Methodological Coherence and Evidence Recognition in Digital Health Systematic Reviews: Sample-based Meta-research Study (e78210)
Uwe Buddrus, Jan-David Liebe. 61

Impact of Patient Portals on Asthma and Chronic Obstructive Pulmonary Disease–Related Quality-of-Care Outcomes: Systematic Review (e83483)
Sukriti Kc, Nicholas Hopkinson, Ana Neves, Azeem Majeed, Anthony Laverty. 80

eHealth Approaches to Multimorbidity Management: Scoping Review of Interventions and Multidisease Strategies (e88021)
Katja Bochtler, Salima Houta. 316

XSL•FO
RenderX

Original Papers

Assessing Patient Discomfort in Smartphone-Based Teledentistry From the Perspective of Dental Professionals: Qualitative Interview Study (e81163)
Margaret Shenouda, Boyen Huang, Emily Schultz, Mohamed Estai, Sarbin Ranjitkar, Jeff Louie, Patimaporn Pungchanchaikul. 115

Online Health Information–Seeking Among Older Adults and Predictors of Use, Motivations, and Barriers in the Context of Healthy Aging: Cross-Sectional Study (e77557)
Yves Bachofner, Alexander Seifert, Samin Sepahniya, Carlo Fabian. 126

Patient Comfort With Smartphone-Based and Intraoral Camera Imaging in Teledentistry: Observational Cross-Sectional Study (e95563)
Martha Lucas, Boyen Huang, Ryan Ip, Angela Hastings, Emily Schultz, Yao-Yi Chiang, Jeffrey Louie. 139

Tweets Surrounding Pharmaceutical Drug Brands With Top Direct-to-Consumer TV-Advertising Budgets: Social Media Listening Study (e85641)
Wendy Macias, María Navarro Castillo. 150

Demonstrating a Social Intelligence Analysis Framework for Loneliness: Infodemiology Approach (e59861)
Hurmat Shah, Mowafa Househ, Loulwah Alsumait, Altaf Alfarhan. 164

Association Between Access to Health Information and Frailty in Older Japanese Adults: Web-Based Cross-Sectional Study (e83642)
Noriko Hori, Jiaqi Li, Yosuke Osuka. 181

Investigating Strategies for Sustaining Modern Health Information Systems for Enhanced Care Delivery in Fragile Resource-Constrained Settings Through Electronic Medical Record Implementation in Haiti: Qualitative Study (e82396)
Benito Isaac, Annie Michaelis, Meredith Casella Jean-Baptiste, Loune Viaud, Wesler Lambert. 190

Global Trends in Phenylketonuria Treatment Research, 2000-2025: Bibliometric Analysis (e90419)
Sitong Yang, Kaichao Song, Qingbo Chen, Jiandong Jiang, Lulu Wang. 203

A Comprehensive Approach to Days' Supply Estimation in a Real-World Prescription Database: Algorithm Development and Validation Study (e83465)
Maria Malk, Kerli Mooses, Marek Oja, Johannes Holm, Hanna Keidong, Nikita Umov, Sirli Tamm, Sulev Reisberg, Jaak Vilo, Raivo Kolde. . .
2 2 6

Knowledge, Attitudes, and Perceptions Regarding Mini Implants Among Orthodontic Practitioners in the Kurdistan Region: Cross-Sectional Study (e86896)
Soran Hameed, Aras Rauf. 240

Comparison of Artificial Intelligence Tools With Human Coding for Sentiment, Topic, and Thematic Analysis Tasks of Public Health Datasets During the COVID-19 Pandemic in Australia: Case Study (e80824)
Danielle Hutchinson, Lauren Lee, Haley Stone, Aye Moa, Holly Seale, C MacIntyre. 252

AI-Based Diagnostic Platform Capabilities With Lyme Disease as a Use Case: Integrative Exploration (e87529)
Sarah Maxwell, Connie McNeely, Abdollah Zeraatpisheh, Chris Brooks, Claire Maxwell, Kevin Thomas. 278

Automating Diagnosis of Skin Neglected Tropical Diseases via Patient Metadata through Machine Learning Model with Adaptive Balancing and Dual Cross-Validation: Retrospective Diagnostic Accuracy Study (e84966)
Yohannes Minyilu, Mohammed Yimer, Million Meshesha. 291

XSL•FO
RenderX

Associations Between Hospital Structural Characteristics and Adoption of Public Health Data Integration and Automation: National Cross-Sectional proofsStudy (e86263)
Hanadi Hamadi, Alaysia Alford, Chloe Smith Lopez, Durron Baker, James Samaniego, Jessica Jin. 335

Associations Between Digital Attitudes, Digital Inequalities, and Health Care Use Among Recipients of a Community-Based Laptop Distribution Program: Cross-Sectional Study (e97872)
Tianca Crocker. 347

Financial Literacy and Financial Toxicity Among US Veterans: Cross-Sectional Survey Informing Public Health Informatics Screening (e97291)
Harman Sabharwal, Ramona Raya. 352

Burden of Stroke in China From 1990 to 2023: Analysis From the Global Burden of Disease Study 2023 (e86225)
Lijuan Fu, Wencai Jiang, Yanhua Peng, Xianjie Zhang, Rui Zhou. 365

Building Enhanced Public Health Data Systems With a Situational Awareness and Learning Tool: Focus Group Study (e77379)
Cole Brokamp, Carson Hartlage, Tiffany Mattingly, Pierce Kuhnell, Andrew Vancil, Andrew Beck, David Hartley. 378

Persona Development in Washington State: Mixed Methods Approach Using Statewide Survey Data (e75422)
Jordan Garcia, Rebecca Hills, Debra Revere, Jaewon Lim, Adam Elder, Chris Baumgartner, Bryant Karras, Janet Baseman. 387

Instagram Memes of Oral Nicotine Pouches: Content Analysis Study (e84025)
Samia Amin, Yunping Wu, Samantha Baptista, Scott Okamoto, Thaddeus Herzog, Crissy Kawamoto, Pallav Pokhrel. 405

Assessment of the Cultural Nuances in COVID-19 Vaccine Uptake Through a Comparative Analysis of English and Spanish Facebook Posts in Tarrant County, Texas: Longitudinal Study (e72465)
Ana Aleksandric, Anisha Dangal, Shirin Nilizadeh, Gabriela Mustata Wilson. 415

Social Media Use, Professionalism, and Professional Identity Among Medical Students and Early-Career Health Care Graduates in Sudan: Cross-Sectional Survey Study (e89260)
Fatima Mohammed, Enas Omer, Ammar Elgadi, Ola Mamoun, Mohamed Alsayed, Saad Elsayed, Ragheda Musaad, Salma Alam, Mohamed Adam, Bayadir Adam, Sara Elshafie, Mohanad Elamin, Omer Mohamed. 433

Topic and Sentiment Trends in Semaglutide Discussions on X: Subpopulation-Based Longitudinal Analysis (e80660)
Parisa Momeni, Gabriel Laverghetta, Jay Ligatti, Lingyao Li. 447

The Use of Tomographs in Brazil's National Health System: Case Study on the Efficiency of the Public Network in Rio Grande do Norte (e83494)
Tiago Barreto, Elinaldo de Oliveira Júnior, Israel dos Santos Felipe, Jordana de Lima Paiva, Nicolas Rodrigues Veras, Lorena de Macedo Silva, João Macedo da Costa, Érika Santos de Aragão, Thaísa Góis Farias de Moura Santos Lima, Célio da Costa Barros, Rafael Silveira e Silva, Fernando Rocha de Andrade, Pablo Holanda Cardoso, Andréa Santos Pinheiro, Guilherme Medeiros Machado, Claudia Miranda Veloso, Luciana Simões Camara Leão, Karilany Dantas Coutinho, Ricardo de Medeiros Valentim. 474

Workflow-Based Information Management Framework for Multicenter Research Studies: Design and Development (e81119)
Hasan Sulaeman, Mars Stone, Roberta Bruhn, Karla Zurita, Anh Nguyen, Vincent Chiang, Jefferson Jones, Xutao Deng, Brian Custer, Michael Busch, Eduard Grebe. 490

The Readability of Patient-Facing Information in Public Reporting System: Cross-Sectional Evaluation of the City of Helsinki Website (e93241)
Lanmei Du, Reima Suomi. 502

XSL•FO
RenderX

Using Immunization Data Systems to Track Zero-Dose Children in Bangladesh: Mixed Methods Exploratory Study (e88951)
Samiha Yunus, Prem Singh, Alan John, Elizabeth Oliveras, Md Uddin, Nurul Alam, Md Ali, Hemel Das, Chris Morgan. 516

Investigation of Community Behaviors, Socioeconomic Factors, and Breakthrough COVID-19 Infections Among Vaccinated Individuals: Cross-Sectional Study (e76679)
Matthew McDonald, Bathri Vajravelu. 527

Fairness Correction in COVID-19 Predictive Models Using Demographic Optimization: Algorithm Development and Validation Study (e78235)
Naman Awasthi, Saad Abrar, Daniel Smolyak, Vanessa Frias-Martinez. 540

Data Access, User Engagement, and Interoperability Standards in Public Health Data Portals: Environmental Scan Study (e95523)
Yogita Chaudhary, Lathika R, Solomon Thirumurugan, Manikandan Kumaraguru, Sharan Murali, Rizwan Abdulkader. 557

Research Letters

Explainable Machine Learning for Public Health Informatics in HEDIS Childhood Immunization Status Combination 10 (e90241)
JiaLi Ruan. 466

AI-Related Stress and Suicidal Tendencies in a Primary Care Cohort: Clinical Audit (e84661)
Shemarah Lindo. 470

XSL•FO
RenderX

Bias in AI in Epic Electronic Medical Records and Its Impact on Executive Communication: Conceptual Analysis

Naomi Pemberton, MPA, CPC

School of Health Sciences, Herbert H. and Grace A. Dow College of Health Professions, Central Michigan University, 1200 S Franklin St, Mount Pleasant, MI, United States

Corresponding Author:
Naomi Pemberton, MPA, CPC
School of Health Sciences, Herbert H. and Grace A. Dow College of Health Professions, Central Michigan University, 1200 S Franklin St, Mount Pleasant, MI, United States

Abstract

Background: AI has become an essential component of modern health care delivery in Epic (Epic Systems Corporation) electronic medical record (EMR) systems, supporting predictive analytics, diagnostic decision-making, and population health management. Despite these advancements, evidence reveals that AI algorithms can perpetuate or even amplify existing health inequities through biased training data and flawed model design. Such algorithmic bias poses ethical challenges for health care leadership, regulatory compliance, and executive communication, especially in ensuring patient equity, transparency, and public accountability.

Objective: This conceptual paper examines how algorithmic bias in Epic's AI modules influences executive decision-making, organizational communication, and trust within health care systems. It integrates organizational communication theory and public health informatics research to propose a framework for ethical, transparent, and equitable communication in AI-integrated health care settings.

Methods: Drawing upon the ethical communication and algorithmic trust framework (ECATF), this paper synthesizes interdisciplinary literature on AI bias, data governance, and leadership communication. The framework explains how transparent executive communication creates stakeholder trust in the context of bias identification and regulatory oversight.

Results (Conceptual Findings): This conceptual analysis suggests that algorithmic bias influences leadership communication, equity framing, and governance strategies in AI-integrated health care systems. Incorporating AI bias auditing alongside regulatory monitoring and public education initiatives may support fairness, accountability, and health literacy across communities.

Conclusions: As AI continues to shape health care leadership and policy, ongoing evaluation, ethical foresight, and regulatory vigilance are essential. Transparency, collaborative governance, and adaptive education will be necessary to ensure that AI supports equitable innovation rather than reinforcing unintended harm.

(Online J Public Health Inform 2026;18:e85584) doi:10.2196/85584

KEYWORDS

AI; health care; algorithmic bias; AI in health care; Epic electronic medical records; executive communication

Introduction

AI technologies are redefining health care operations, driving data-informed decision-making, and transforming how patient care is delivered. In Epic (Epic Systems Corporation) electronic medical record (EMR) systems, AI models are widely used for tasks such as predicting readmission risks, identifying care gaps, and optimizing clinical workflows. Examples include Epic's predictive analytics and risk stratification tools, which are used to identify high-risk patients, monitor sepsis indicators, and support clinical decision-making. Although these tools promise efficiency, reduced administrative burden, and better patient outcomes, their use in health care systems has also raised concerns regarding bias, fairness, and ethical governance [1,2].

Algorithmic bias stemming from incomplete, unrepresentative, or historically biased data usually leads to unequal treatment recommendations, inaccurate risk stratifications, and disadvantages for marginalized populations. Such biases disproportionately affect racial and socioeconomic minorities, amplifying existing health disparities [3,4]. These concerns extend beyond technical accuracy because they influence how executives interpret AI-driven data, communicate organizational strategies, and shape institutional trust [5].

Health care leaders increasingly depend on AI-generated insights for strategic planning, performance evaluation, and policy formulation [6]. Yet, without a contextual understanding of data provenance and algorithmic limitations, executive communication may unintentionally reinforce inequities [7]. As AI systems become more embedded in executive

XSL•FO
RenderX

decision-making, concerns regarding transparency, accountability, and trust become more urgent.

This paper examines the ways AI bias in Epic's EMR systems influences executive communication and ethical decision-making practices. By integrating organizational communication theory, it explores how leadership framing, transparency, and relational trust can mitigate bias-related harm in health care organizations. Ultimately, this conceptual analysis underscores the need for equitable AI governance and communicative integrity in data-driven health care leadership.

Methods

This conceptual analysis used a narrative literature synthesis to examine the relationship between algorithmic bias, executive communication, and ethical governance in AI-integrated health care systems. Literature published between 2018 and 2025 was identified through PubMed, Scopus, Google Scholar, and IEEE Xplore using combinations of keywords including "AI bias," "Epic EMR," "algorithmic trust," "executive communication," "healthcare AI ethics," and "algorithmic transparency."

Articles were included if they addressed health care AI, algorithmic fairness, organizational communication, trust in AI systems, or ethical governance in clinical or administrative health care settings. Priority was given to peer-reviewed journal articles and interdisciplinary scholarship addressing health care leadership and AI governance.

The selected literature was synthesized conceptually to identify recurring themes related to algorithmic bias, executive communication, transparency, and stakeholder trust. These themes informed the development and application of the ethical communication and algorithmic trust framework (ECATF).

Results (Conceptual Findings)

The ECATF serves as the theoretical foundation for this paper because it links organizational communication theory with contemporary health informatics. Rooted in the principles of message framing, relational trust, and transparency, the ECATF explains how health care executives interpret, communicate, and act upon AI-generated insights in complex institutional ecosystems. The framework also highlights how communication mediates the relationship between algorithmic bias and stakeholder trust while emphasizing leadership's responsibility to promote transparency and accountability in data-driven decision-making.

The framework consists of 4 interconnected components. In the ECATF, algorithmic bias may be operationalized through measures such as demographic disparities in prediction accuracy, false-positive or false-negative rates across patient populations, and documented bias audit outcomes. Executive communication transparency may be assessed through disclosure frequency, message clarity, stakeholder engagement, leadership briefings, and public reporting practices directed toward clinicians, patients, regulators, and community stakeholders. Trust may be evaluated through employee confidence surveys, patient trust perception measures, and stakeholder feedback regarding institutional transparency. These dimensions help explain how the ECATF functions across organizational communication and AI governance contexts [7].

The first stage is algorithmic bias identification. This stage involves recognizing and evaluating inequities embedded in Epic's algorithms. It requires an understanding of how data inputs, training sets, and model parameters may inadvertently disadvantage certain populations. Effective bias identification requires collaboration among clinicians, data scientists, and communication professionals to address the technical and social dimensions of bias. The second stage is the executive communication response. Once bias is identified, leadership must determine how to frame and communicate these findings. Communication framing theory suggests that the way leaders communicate bias-related findings can either increase skepticism or strengthen stakeholder trust. When executives communicate openly about algorithmic limitations, acknowledge uncertainty, and explain corrective actions clearly, they are more likely to maintain credibility and preserve organizational trust. The third stage involves trust mediation. Trust operates as the central component of the ECATF. Communication practices shape both internal trust (among clinicians and staff) and external trust (among patients and the broader community). When leaders engage in open dialogue about algorithmic performance, limitations, and safeguards, they foster psychological safety and strengthen the institution's integrity. Conversely, opacity or defensive messaging can amplify skepticism and weaken stakeholder relationships. The fourth stage is the equity feedback loop. This final component establishes ongoing mechanisms for transparency, bias correction, and stakeholder input. This feedback loop ensures that communication is not merely reactive but iterative, integrating lessons from audits, community feedback, and performance evaluations. The loop encourages continuous ethical reflection and adjustment of AI governance practices.

In this model, executive communication functions as a moderator and intervention point between AI bias and public perception. The ECATF thus argues that ethical communication is central to responsible AI governance. By institutionalizing transparent dialogue, consistent messaging, and equitable engagement, leaders can use communication to strengthen organizational trust and support equitable governance.

Discussion

Overview

This conceptual analysis highlights how algorithmic bias shapes leadership communication, policy responses, and equity-related decision-making. Bias in AI systems manifests through structural inequities, incomplete datasets, and technical limitations in algorithmic design. In Epic's EMR platform, predictive models used for patient risk scoring, readmission forecasting, and clinical prioritization may yield skewed or inaccurate outputs across racial, gender, or socioeconomic groups [3,4]. These discrepancies reflect inequities embedded in the data infrastructure itself [3,7]. When biased AI outputs inform administrative decisions in areas such as funding

XSL•FO
RenderX

allocation, staffing models, or quality improvement metrics, they can inadvertently perpetuate disparities in care delivery.

Health care executives who rely on AI-generated analytics to inform strategic and operational decisions often face the challenge of interpreting complex datasets without specialized expertise in machine learning or data ethics. Such a situation causes vulnerability, as algorithmic bias may go unrecognized or be misinterpreted. Consequently, effective executive communication is essential to bridge the divide between technical accuracy, ethical responsibility, and stakeholder understanding.

Recent scholarship, including the work of Lacmanovic and Skare [8], emphasizes the importance of systematic AI bias–auditing frameworks to promote fairness, transparency, and accountability in algorithmic systems. Their research highlights continuous model evaluation, stakeholder collaboration, and transparent reporting to support the development of equitable AI systems over time. Embedding such auditing mechanisms in Epic's predictive models can enhance not only algorithmic reliability but also the clarity and credibility of leadership communication.

From a communication theory perspective, frameworks such as framing theory and relational trust theory help explain how health care leaders communicate uncertainty, limitations, and the steps they are taking to correct bias. Relational trust theory, meanwhile, suggests that open and empathic communication builds psychological safety in organizations and fosters credibility among external stakeholders [9]. These communication approaches enable executives to position AI as a tool that supports equitable and transparent health care practices.

Equally important is community education and engagement. Communication should extend beyond the institutional hierarchy to reach the patients and families who are most affected by AI-driven health care decisions. Community forums, digital literacy campaigns, and participatory governance models can improve public understanding of how AI-integrated health systems can enhance accountability and reinforce equity-centered leadership.

Expanding the Conceptual Framework Beyond Epic

Although Epic is among the most widely used EMR systems globally, algorithmic bias is not unique to its platform. Similar ethical and communicative challenges exist across other digital health systems, including Oracle Health (formerly Cerner; Oracle Corporation) and Google Health AI (Google LLC). These systems rely on predictive modeling and large-scale data analytics to optimize care coordination, yet each faces comparable issues of transparency, model generalizability, and stakeholder trust.

The ECATF provides a framework for evaluating these challenges across diverse systems. For instance, Oracle Health's clinical decision-support algorithms, which are used to flag deteriorating patients, have exhibited performance variation across demographic subgroups, prompting concerns about algorithmic equity. Likewise, Google Health AI's diagnostic imaging models, despite demonstrating high technical accuracy, have faced scrutiny regarding data diversity and model explanation.

Applying the ECATF across these platforms allows leaders to adopt consistent communication protocols: explicitly identifying algorithmic limitations, explaining model updates, and inviting public and professional feedback. In doing so, organizations move beyond platform-specific problem-solving toward a unified model of ethical AI governance. The framework thus serves as a blueprint for strengthening transparency and trust across the broader digital health ecosystem, ensuring that technology development and leadership communication evolve in tandem toward shared equity goals.

Practical Implications

This conceptual analysis highlights that bias auditing and ethical communication should function together in health care leadership. Executives must not only understand the technical foundations of AI but also develop competencies in transparent, audience-sensitive communication. Implementing periodic algorithmic audits, as recommended by Lacmanovic and Skare [8], ensures that model performance is continually assessed for fairness and accuracy. These audits should be accompanied by communication strategies that disclose audit findings and describe institutional responses.

From an operational standpoint, transparent reporting builds institutional credibility, fosters regulatory compliance, and strengthens interdepartmental collaboration [10]. Clear communication also plays a preventive role by reducing misinformation and improving alignment around accountability and inclusion.

At the community level, inclusive communication practices are equally vital. Public education campaigns that explain how algorithms influence care decisions and how biases are identified and corrected promote health literacy and trust [9,10]. Engaging clinicians, patient advocacy groups, and local communities in discussions about bias auditing democratizes and reinforces public engagement in AI governance. Participatory dialogue can help organizations maintain ethical oversight and human-centered AI governance.

Limitations and Future Directions

As a conceptual paper, this study does not include empirical data or hypothesis testing. Instead, it establishes a theoretical foundation linking organizational communication with health informatics to address AI bias. Future research should adopt mixed methods designs, combining quantitative measures of trust and equity perception with qualitative interviews exploring executive communication strategies. Experimental and simulation studies could further evaluate how leadership framing influences clinician confidence and patient trust in AI-assisted care.

Longitudinal research can help assess how bias-auditing programs influence organizational transparency, equity outcomes, and public perception. Moreover, developing standardized metrics for communication effectiveness, such as message clarity, stakeholder engagement, and perceived fairness, would help operationalize the ECATF model across diverse

XSL•FO
RenderX

health care contexts. Such empirical validation will enable health systems to translate ethical principles into more consistent communication and governance practices.

Conclusions

AI has significant potential to improve health care efficiency, precision, and patient outcomes, yet it simultaneously presents complex ethical and communicative challenges. Algorithmic bias, if left unaddressed, can weaken trust and perpetuate inequity. The ECATF framework underscores transparent communication as the mechanism through which organizations sustain credibility and moral legitimacy.

To achieve ethical AI adoption, health care leaders must integrate bias auditing, data literacy training, and inclusive community engagement. These actions promote equity, public trust, and transparent governance in data-driven health systems. However, the long-term effects of algorithmic bias in health care are uncertain. As AI systems continue to evolve and influence clinical decision-making, the full scope of their impact, whether beneficial or detrimental, has yet to be evaluated.

While algorithmic tools can potentially enhance diagnostic accuracy, streamline operations, and personalize care, unchecked biases may also reinforce existing disparities or introduce new forms of inequity. Their effects can manifest over time, shaping care delivery patterns, patient outcomes, and institutional trust in ways that are difficult to predict or reverse.

Therefore, ongoing evaluation, ethical foresight, and regulatory review are necessary to understand the relationship between AI systems and human judgment. Future research should prioritize longitudinal assessments of AI bias and its influence on patient outcomes, clinical decision-making, and leadership communication. Transparency, collaborative governance, and adaptive education ensure that AI remains a force for positive transformation rather than a mechanism of inadvertent harm in health care.

Through sustained transparency, collaborative learning, and community partnership, health care organizations can ensure that AI remains an instrument of equity and ethical progress in an increasingly data-driven world.

Acknowledgments

The author gratefully acknowledges Dr Beth Taylor, Dr Mark Taylor, and Dr Naomi Deshore for their unwavering support, encouragement, and mentorship. Their guidance and commitment to academic excellence have been deeply appreciated, and the author is sincerely grateful for their contributions throughout this scholarly journey.

During the preparation of this manuscript, the author used OpenAI's ChatGPT to assist with language editing, grammar, sentence structure, and organization of the manuscript. All scientific content, literature selection, interpretation of the findings, conclusions, and final editorial decisions were performed solely by the author. The author reviewed, verified, and takes full responsibility for the accuracy and integrity of the manuscript.

Funding

The author declared no financial support was received for this work.

Conflicts of Interest

None declared.

References

1. Meskó B, Hetényi G, Győrffy Z. Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv Res 2018 Jul 13;18(1):545. [doi: 10.1186/s12913-018-3359-4] [Medline: 30001717]
2. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc 2020 May 30;2020(12):191-200. [Medline: 32477638]
3. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
4. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018 Dec 18;169(12):866-872. [doi: 10.7326/M18-1990] [Medline: 30508424]
5. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019 Dec 1;46(4):629-650. [doi: 10.1093/jcr/ucz013]
6. Theodorakopoulos L, Theodoropoulou A, Halkiopoulos C. Cognitive bias mitigation in executive decision-making: a data-driven approach integrating big data analytics, AI, and explainable systems. Electronics (Basel) 2025;14(19):3930. [doi: 10.3390/electronics14193930]
7. Kordzadeh N, Ghasemaghaei M. Algorithmic bias: review, synthesis, and future research directions. Eur J Inf Syst 2022 May 4;31(3):388-409. [doi: 10.1080/0960085X.2021.1927212]
8. Lacmanovic S, Skare M. Artificial intelligence bias auditing—current approaches, challenges and lessons from practice. Rev Account Finance 2025 May 19;24(3):375-400. [doi: 10.1108/RAF-01-2025-0006]

XSL•FO
RenderX

9. Park K, Yoon HY. AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI. Humanit Soc Sci Commun 2025;12(1):1-3. [doi: 10.1057/s41599-025-05116-z]
10. Ratwani RM, Sutton K, Galarraga JE. Addressing AI algorithmic bias in health care. JAMA 2024 Oct 1;332(13):1051-1052. [doi: 10.1001/jama.2024.13486] [Medline: 39230911]

Abbreviations

ECATF: ethical communication and algorithmic trust framework
EMR: electronic medical record

Edited by E Mensah; submitted 09.Oct.2025; peer-reviewed by A gopinath, J Bush, N Deshore, S Mohanadas; revised version received 19.May.2026; accepted 05.Jun.2026; published 31.Jul.2026.

Please cite as:
Pemberton N
Bias in AI in Epic Electronic Medical Records and Its Impact on Executive Communication: Conceptual Analysis
Online J Public Health Inform 2026;18:e85584
URL: https://ojphi.jmir.org/2026/1/e85584
doi:10.2196/85584

© Naomi Pemberton. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 31.Jul.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Transforming Pediatric Care Through AI: Bridging the Digital Divide in Health Informatics

Mercy Mbogori-Kairichi, MSc; R Sterling Snead, MS; Julia Marczika, BS; Radha Ambalavanan, MPhil; Gideon Towett, MSc; Alex Malioukis, MSc

Research Department, The Self Research Institute, 18452 E 111th St, Broken Arrow, OK, United States

Corresponding Author:
Mercy Mbogori-Kairichi, MSc
Research Department, The Self Research Institute, 18452 E 111th St, Broken Arrow, OK, United States

Abstract

Health informatics and artificial intelligence (AI) technologies are increasingly influencing pediatric health care delivery across diverse health system contexts. These technologies offer opportunities to improve diagnostic accuracy, personalized treatment approaches, and access to care globally. This viewpoint examines how health and public health informatics frameworks, when integrated with AI technologies, may help address persistent challenges in global pediatric care delivery. This paper is a viewpoint informed by selected published studies and international digital health guidance rather than a systematic review. Evidence from clinical implementations suggests that AI applications embedded in standardized electronic health records can facilitate improved pediatric diagnostic processes. For instance, machine learning–based algorithms to diagnose serious bacterial infections among febrile infants have shown high diagnostic accuracy and reduced unnecessary invasive procedures in certain clinical contexts. Case studies from the Pediatric Emergency Care Applied Research Network decision rules, neonatal intensive care units, and autism screening programs reflect diverse applications of AI-enabled clinical decision support across pediatric settings. However, there are concerns regarding implementation due to limitations in interoperability of health information systems, gaps in data standardization, inadequate digital infrastructure in resource-limited settings, and issues related to algorithmic bias and equitable access. We argue that strategic development of interoperable health information systems, standardized data governance frameworks, and equitable digital infrastructure is essential to responsibly realize the potential of AI-enhanced pediatric care at scale.

(Online J Public Health Inform 2026;18:e82108) doi:10.2196/82108

KEYWORDS

health informatics; digital health; pediatrics; artificial intelligence; health information systems; electronic health records

Introduction

Artificial intelligence (AI) is increasingly being integrated into health care delivery and digital health systems, contributing to the modernization of medical practice [1]. Health informatics provides the foundational infrastructure and standardized data models that enable AI algorithms to function effectively in real-world settings. The COVID-19 pandemic has increased the use of digital health worldwide, accelerating the integration of AI into pediatric care systems [2]. Emerging evidence suggests promising applications in pediatric settings. For example, machine learning (ML) algorithms applied to clinical data from febrile infants have demonstrated high diagnostic performance, achieving 98.6% sensitivity and 74.9% specificity in identifying serious bacterial infections. The random forest model showed superior diagnostic accuracy (area under the receiver operating characteristic curve: 0.96) and was estimated to potentially reduce 68.5% of unnecessary lumbar punctures [3]. These findings suggest that AI-driven diagnostic models, when supported by robust health informatics infrastructures and structured electronic health record (EHR) environments, can enhance pediatric diagnostic accuracy while reducing unnecessary interventions.

Despite some progress, pediatric health informatics has lagged behind adult-focused digital health advancements. Researchers have raised concerns about the validity and reliability of applying adult-trained AI models to pediatric populations, with potential implications for care quality [4]. In low- and middle-income countries (LMICs), the synergy between AI and informatics may provide a valuable platform for enhancing access and quality in pediatric care. According to the World Health Organization (WHO), under-five mortality remains a major global health challenge, with millions of deaths occurring annually from largely preventable causes [5]. In this context, AI-enabled decision support systems have the potential to facilitate early diagnosis and improved management pathways, although real-world impact remains uneven. However, substantial barriers to adoption persist, including interoperability challenges, data standardization gaps, digital infrastructure limitations, and equity concerns in resource-constrained settings. Much of the current literature examines AI applications and health informatics infrastructure separately, with comparatively

XSL•FO
RenderX

limited attention to their integrated implementation within pediatric care settings, especially in relation to equity and real-world challenges. This gap is critical because pediatric populations have unique physiological, developmental, and care delivery characteristics that require specialized informatics frameworks and AI models distinct from adult-focused approaches.

This viewpoint examines how public health informatics frameworks, when integrated with AI technologies such as ML, natural language processing (NLP), and predictive analytics, can support efforts to improve pediatric care delivery across diverse health system contexts. We review successful applications of health informatics and AI in pediatric settings, identify persistent challenges hindering integration, and discuss considerations for implementing informatics frameworks that responsibly leverage AI to support pediatric care. This analysis is intended for pediatric health care providers, health informaticists, policymakers, and researchers working to advance digital health solutions in pediatric care.

Evidence Informing This Viewpoint

This viewpoint is informed by a purposive selection of published literature and international guidance reports. We conducted a targeted search of PubMed and Google Scholar using key terms related to pediatrics, AI, health informatics, and digital health. We also used gray literature sources, including policy and guidance reports from the WHO and World Bank, to identify relevant evidence published between 2014 and 2025. Search concepts included combinations of terms, such as pediatrics, children, AI, ML, health informatics, digital health, EHRs, clinical decision support, telemedicine, implementation, and health equity. Using the selected literature, we identified 3 illustrative case examples to illustrate successful applications across distinct domains. These include clinical decision support (Pediatric Emergency Care Applied Research Network [PECARN] head trauma and abdominal trauma rules), predictive modeling (neonatal sepsis detection and bronchopulmonary dysplasia [BPD] prediction), and automated screening (autism spectrum disorder [ASD]). These cases were chosen to represent diverse pediatric age groups, clinical contexts, and informatics applications.

This methodology had several limitations. This synthesis may preferentially reflect selection bias toward well-documented implementations in high-resource settings. Emerging applications and ongoing trials may not be fully captured. The evidence base for implementation in LMICs remains limited, representing an important research gap highlighted in this viewpoint. Textbox 1 summarizes the key potential benefits and implementation challenges of AI and health informatics in pediatric care, organized by domain. Figure 1 presents the conceptual framework informing this viewpoint, illustrating conceptual linkages from health informatics infrastructure through AI applications to pediatric care outcomes.

Textbox 1. High-level synthesis of key benefits and implementation challenges of artificial intelligence and health informatics in pediatric care. Benefits reflect emerging evidence from selected implementations and should be interpreted cautiously. Challenges highlight issues that require addressing.

Potential benefits

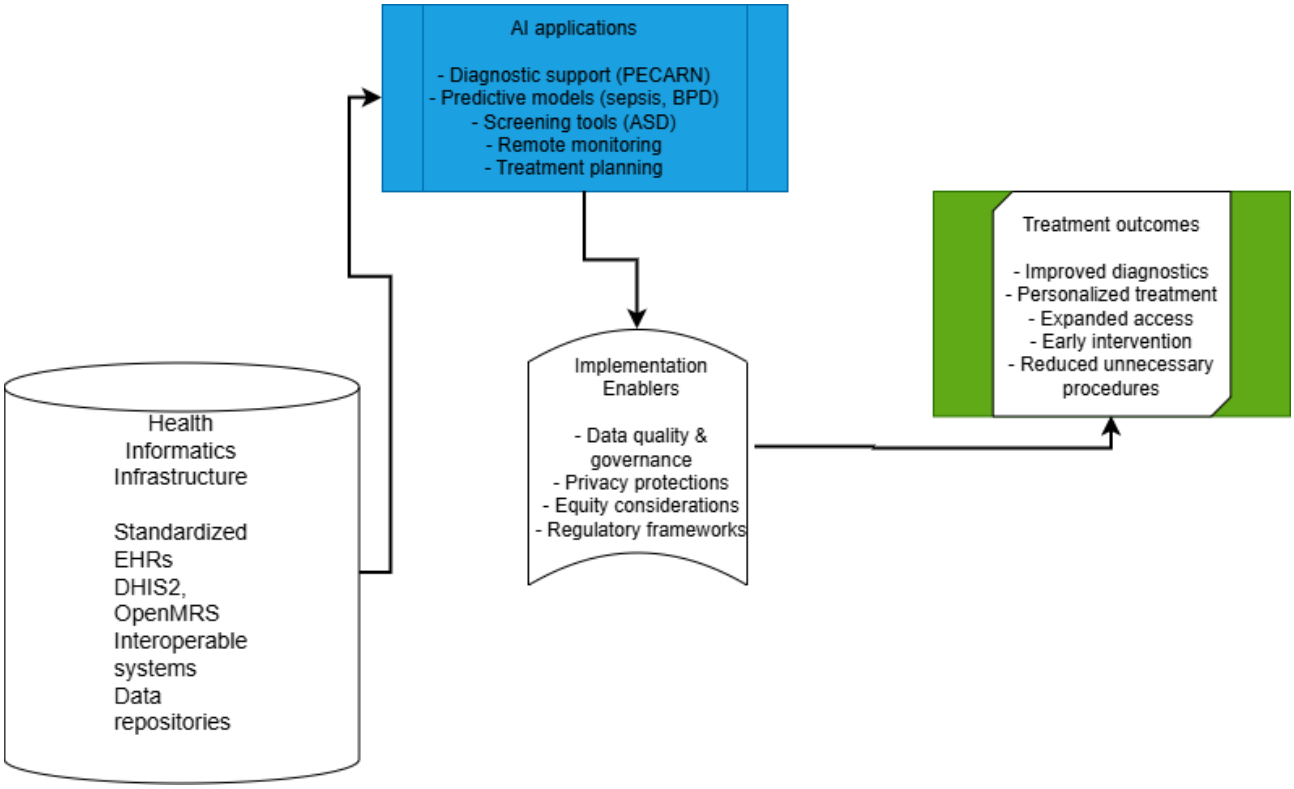
- Diagnostic support—artificial intelligence (AI) tools may improve diagnostic accuracy for conditions such as pneumonia
- Predictive modeling—AI models show promise for early detection of high-risk conditions, including neonatal sepsis and bronchopulmonary dysplasia
- Personalized treatment—AI-supported clinical decision tools may inform individualized treatment selection
- Accessibility—telemedicine platforms integrated with AI may expand access to pediatric specialists in underserved areas
- Efficiency—automated screening tools and remote monitoring systems enable earlier intervention
- Remote monitoring—mobile health platforms with AI analytics may support chronic disease management

Implementation challenges

- Data quality—limited availability of high-quality, representative pediatric data sets
- System integration—incompatibility between legacy health information systems and newer AI technologies
- Algorithmic bias—risk of perpetuating health inequities when models are trained on unrepresentative data
- Privacy and ethics—need for robust privacy protections and informed consent frameworks
- Regulatory variation—limited pediatric-specific AI governance
- Socioeconomic disparities—unequal distribution of AI implementations across income levels

XSL•FO
RenderX

Figure 1. Conceptual framework for AI-enhanced pediatric care delivery. This framework illustrates the pathway from health informatics infrastructure through AI applications to improved pediatric care outcomes as explored in this viewpoint. The arrows represent conceptual relationships. AI: artificial intelligence; ASD: autism spectrum disorder; BPD: bronchopulmonary dysplasia; DHIS2: District Health Information Software 2; EHR: electronic health record; OpenMRS: Open Medical Record System; PECARN: Pediatric Emergency Care Applied Research Network.



Population-Level Benefits of Informatics and AI in Pediatric Health

Health informatics and AI offer several potential benefits that would improve global pediatrics. These technologies may improve accuracy, personalization of pediatric care, accessibility, early detection and intervention, and support for pediatric mental health services. One such benefit is efficiency and accuracy. Algorithm-based diagnostic tools have demonstrated effectiveness in selected clinical contexts. For instance, convolutional neural networks have been applied to pediatric medical imaging, including the diagnosis of leukemia [6], as well as other conditions such as pneumonia and retinopathy of prematurity [7]. Systematic reviews suggest that some AI systems have performed comparably to human experts in interpreting radiographs and computed tomography (CT) images [8], while validation studies in ophthalmic imaging have demonstrated expert-level performance in specific diagnostic tasks [9]. This evidence suggests that AI can enhance aspects of pediatric care delivery in specific contexts. Beyond diagnostic imaging, AI-driven systems have been explored for continuous monitoring and personalized management in pediatric chronic conditions, including the development of knowledge-enabled conversational systems to support asthma self-management by integrating patient-reported data, environmental information, and domain knowledge to deliver contextualized and personalized guidance [10]. In more acute settings, ML approaches have been applied to pediatric critical care EHR data to identify prognostic clusters and clinically relevant patterns, highlighting the potential of data-driven models in critically ill pediatric populations [11].

Another notable benefit is personalization. Health informatics and AI support the development of targeted treatments. AI analyzes patient characteristics to inform treatment decisions for different pediatric patients. This is an evolution from the usual use of a definite general scheme of therapy. ML algorithms play important roles in such personalization. In pediatric oncology, AI-based decision-making tools are increasingly being developed to support clinicians in selecting a chemotherapy regimen based on tumor biomarkers and patient characteristics [12]. By enabling more precise treatment selection, these tools may reduce adverse effects. AI-driven conversational systems, such as the chatbot Tess, have been implemented to provide adolescents managing chronic conditions with tailored information and emotional support [13]. Digital tools, therefore, show promise in addressing the psychosocial needs of children and adolescents with chronic illnesses.

There is also the benefit of accessibility. Ensuring equitable access to high-quality services remains a persistent challenge in resource-constrained settings, where limitations in infrastructure, workforce capacity, and system-level quality standards affect service delivery [14]. Telemedicine platforms supported by health informatics technology play an important role in improving accessibility. By incorporating NLP and real-time analyses, these platforms enable patients, especially children in underserved areas, to seek consultations from specialists. AI-powered translation and NLP tools have been

XSL•FO
RenderX

explored to address communication and language barriers in global health settings, which may be particularly relevant for pediatric populations in multilingual and resource-constrained contexts [15].

The benefit of early detection and intervention cannot be overlooked. Health informatics and AI applications in pediatrics extend to opportunities for early detection and intervention. For instance, AI-enhanced imaging modalities in pediatric cardiology, including echocardiography and other cardiac imaging techniques, have demonstrated potential to improve diagnostic insight and support timely detection of congenital heart disease [16]. For chest radiography, AI applications in tuberculosis and pneumonia diagnosis have demonstrated promising diagnostic performance in pediatric populations, with potential to support clinical decision-making across diverse care settings [8]. Apart from diagnostics, health informatics and AI expand the availability of preventive care. For example, health informatics and AI have been used to predict sepsis and BPD in neonates in the neonatal intensive care unit [17], while other ML approaches have explored early identification of neonatal sepsis [18]. Such predictive models may support earlier clinical intervention when interpreted within established care pathways.

Critical Evidence: Real-World AI Applications Transforming Pediatric Care

Health informatics and AI are not merely concepts in pediatric health care, as these technologies have demonstrated measurable outcomes in clinical settings. The following case studies illustrate applications across 3 domains: clinical decision support, predictive modeling, and automated screening. In this viewpoint, these domains are selected to represent diverse pediatric contexts and informatics approaches, encompassing the PECARN rules, AI in neonatal sepsis, and AI diagnosis of ASD.

PECARN Clinical Decision Rules

The PECARN clinical decision rules represent a large-scale application of evidence-based clinical prediction modeling in pediatric emergency medicine rather than a traditional AI-driven system. PECARN is a coordinated network of pediatric emergency departments across the United States and Canada. Researchers derived and prospectively validated statistical prediction rules to identify children at very low risk of clinically important traumatic brain injury. These predictors were incorporated into a validated clinical decision support rule that assists physicians in identifying children at very low risk for clinically important traumatic brain injury, thereby helping determine when CT imaging can be safely avoided [19]. As one of the earliest large-scale validated prediction frameworks in pediatric emergency medicine, PECARN laid important groundwork for subsequent AI-enabled decision support systems.

A large multicenter prospective validation study involving 7542 children with blunt abdominal trauma and 19,999 children with minor head trauma demonstrated high diagnostic accuracy of the PECARN prediction rules. The intra-abdominal injury rule showed 100% sensitivity for identifying children requiring acute intervention. For traumatic brain injury, the rule demonstrated 100% sensitivity in children younger than 2 years and 98.8% sensitivity in children aged 2 years and older for clinically important traumatic brain injuries. Implementation of these validated clinical prediction rules supports safer decision-making regarding CT imaging and may reduce unnecessary radiation exposure while maintaining high-quality care [19]. This example demonstrates how rigorously validated, data-driven clinical decision support tools, when embedded within structured informatics frameworks, can improve diagnostic accuracy while reducing unnecessary radiation exposure.

ML for Neonatal Condition Prediction

Neonatal sepsis continues to be among the most morbid and deadly illnesses in infants, especially in LMICs. A global systematic review and meta-analysis reported an estimated incidence of 2824 cases per 100,000 live births, with a mortality rate of 17.6%, underscoring the persistent burden of disease worldwide [20]. Early identification is especially challenging in the neonatal intensive care unit, where clinical presentations are often nonspecific and may delay diagnosis and treatment.

Early sepsis identification is critical, yet clinical presentations are often nonspecific. This leads to delayed diagnosis and treatment. One model was trained on large vital sign datasets, including heart rate variability, respiration, and oxygen saturation from high-risk neonates [18]. Such approaches may support earlier clinical intervention and improve outcomes in vulnerable neonatal populations.

Similarly, predictive modeling has been applied to BPD risk assessment in preterm infants. The study showed that using AI in predicting BPD in preterm infants achieved 88% sensitivity and 91% specificity. The study also enrolled 61 preterm infants (gestational age 24‑31 wk) and identified the 26 (43%) infants who developed BPD. When using clinical data alone (birth weight, gestational age, and surfactant treatment), the model showed 74% sensitivity and 82% specificity. Performance improved to 88% sensitivity and 91% specificity when combined with AI-analyzed spectral data from physiological signals [18]. These findings suggest that ML-enhanced analysis of continuous physiologic data may improve early risk stratification in high-risk neonates.

ML for Autism Screening

ASD is a neurodevelopmental condition characterized by differences in social communication and restricted or repetitive behaviors [21]. Early ASD identification and referral to intervention programs are associated with improved developmental outcomes. However, comprehensive diagnostic assessments for ASD require specialized clinical expertise that may be limited in many developing countries. The traditional Autism Diagnostic Observation Schedule-Generic (ADOS) assessment takes between 30 and 60 minutes to administer. Families may wait as long as 13 months between initial screening and diagnosis. ML approaches have been explored to streamline these assessment challenges.

Wall and colleagues at Harvard Medical School developed an AI-based screening tool, which analyzes data from behavioral and cognitive assessments to identify patterns of ASD. Their

XSL•FO
RenderX

study analyzed 612 individuals with autism using existing ADOS datasets, alongside comparison groups from independent cohorts. These included the Autism Genetic Resource Exchange and the Boston Autism Consortium. The research tested 16 different ML algorithms to identify the most effective approach for autism classification. The resulting classifier used only 8 items from the original 29-item ADOS module. This represented a 72.4% reduction in assessment items. This approach achieved an accuracy of 100% sensitivity and 100% specificity in the initial training set [22]. It correctly classified all 612 individuals with autism and all 15 nonspectrum individuals. When tested on independent datasets, the classifier maintained performance with 99.7% accuracy on the Simons Simplex Collection and 100% accuracy on the Boston Autism Consortium dataset [18,22]. These results are promising, but the small sample sizes and need for validation in diverse clinical populations limit generalizability.

Current Applications

Health informatics provides the foundational infrastructure for integrating AI into clinical workflows. Standardized health information systems enable mechanisms for diagnostics, treatment planning, decision support, and patient monitoring. Examples of diagnostic innovations in health care include molecular detection platforms such as M gene–targeted quantitative reverse transcription polymerase chain reaction assays for pathogen identification [23]. The success of health informatics–enabled AI implementations in this area informs the potential for leveraging ML and AI to improve pediatric care. In diagnostic applications, standardized EHR systems organize clinical data that ML algorithms can analyze to support diagnostic decisions. Several studies report that ML systems demonstrate promising diagnostic performance for detecting pulmonary conditions in pediatric chest radiographs [6,8]. Similarly, AI-assisted histopathology has been explored for detecting medulloblastoma. A study by Attallah reported improved diagnostic accuracy for tumor subtyping compared to conventional methods [24]. These AI-supported tools may help reduce diagnostic errors in pediatric practice.

Applications in treatment personalization integrate EHR data with ML to support individualized treatment decisions. Clinical data from EHRs, including genetic and clinical data and patient environmental factors, can inform ML models to predict treatment response. Pediatric oncology has been a particular focus for such applications, where AI supports clinicians in selecting chemotherapy regimens based on the tumor characteristics and patient factors [12]. AI applications have also been explored in surgical planning, where patient-specific anatomic models may assist pediatric surgeons in virtual surgical planning, potentially improving procedural accuracy.

Remote monitoring applications leverage mobile health platforms and wearable devices for pediatric chronic disease management. These systems collect and analyze physiological data from wearables and EHRs for continuous vital sign monitoring among children with chronic conditions. Real-time alerts may support early interventions that could potentially reduce hospital readmission rates [25]. Studies suggest that ML applied to pediatric critical care data may provide prognostic insights and support risk stratification in hospitalized children [11].

Challenges in AI-Enabled Pediatric Systems

Implementation of AI in health care, including pediatric health informatics, faces several interconnected challenges. These challenges include technical challenges related to data quality, system integration, and ethical and legal concerns regarding privacy and transparency, and socioeconomic disparities in access and resources [26]. Technical challenges in pediatric data collection include small sample sizes, ethical constraints on research involving children, and developmental heterogeneity across the pediatric population [27]. AI models trained on historical data may be limited by data that are inadequate, contradictory, or biased. Training datasets often lack adequate representation of diverse populations, including racial and other demographic groups, potentially leading to biased model outputs [28]. High-quality pediatric data from LMICs remains limited due to health system infrastructure gaps and fragmented health information systems [29]. Addressing these gaps requires investment in strong data-sharing platforms and strengthened digital infrastructure to enable representative pediatric data collection globally.

System integration challenges may arise from differences in data structures, data quality, and governance frameworks between existing health information systems and newer AI technologies [30]. Interoperability standards such as Health Level Seven Fast Healthcare Interoperability Resources are being adopted to address these technical barriers, though implementation remains inconsistent globally [31].

There is also a challenge of algorithmic bias that poses a significant problem when deploying health informatics and AI in pediatric care. Models trained on biased datasets may perpetuate existing health inequalities. For example, algorithms developed using data collected in high-income countries may perform poorly in LMICs because epidemiologic profiles and models of health care delivery are different [32,33]. Bias may also arise from subjective labeling of training data or underrepresentation of minority populations in data collection.

Ethical and legal concerns include data privacy protections, regulatory variations, and algorithmic transparency that affect trust in AI systems. Pediatric health data are particularly sensitive, as privacy breaches may have long-term implications for individuals [30]. Most jurisdictions have data protection frameworks. For example, the General Data Protection Regulation in Europe and the Health Insurance Portability and Accountability Act in the United States [34]. Beyond privacy, algorithmic transparency presents major ethical challenges. A review of AI tools in LMIC health care settings noted that many systems relied on complex or “black-box” algorithms with limited interpretability, raising concerns about transparency and clinical trust [32].

Technology changes faster than regulatory policies. The regulatory framework in pediatrics reflects this dynamic. Some

XSL•FO
RenderX

developed nations have advanced AI governance frameworks. For example, Singapore has developed a Model AI Governance Framework. Other nations, such as the United Kingdom and the United States, are continuously developing provisions to regulate AI. Many LMICs appear to be in the earlier stages of developing robust regulatory frameworks for AI in health care [30]. International organizations, including the WHO and International Telecommunication Union, are working to establish global standards and guidance for responsible AI deployment in health [35].

Socioeconomic disparities in AI adoption for pediatric health reflect broader global health inequities. A review by Ciecierski-Holmes et al [32] found that 80% of AI implementations were in upper-middle-income countries. There were only 10% in low-income countries and 10% in lower-middle-income countries. These findings reflect the uneven distribution of AI health care technologies globally. Multiple factors contribute to these disparities, including limited digital infrastructure, a shortage of technical expertise, competing health system priorities, and insufficient investment in health information systems [35]. Advancing equitable AI adoption requires addressing these structural barriers alongside technology development.

Solutions for Scalable Pediatric AI Adoption

Enhanced Data Quality and System Integration

Improving data quality and availability is a critical priority. This is significant because only 50% of AI health care applications in LMICs documented their training datasets adequately [32]. ML performance depends on the quality and representativeness of training data. Given challenges in pediatric research data collection, the successful adoption of AI in pediatrics would largely benefit from collaborative efforts to develop a uniform data source. Such a source would hold pediatric practice data from diverse geographic and demographic contexts. This would create uniform data repositories that provide adequate and high-quality data for the development of health information systems and AI models in pediatric care.

System integration approaches should build on existing health information infrastructure rather than requiring wholesale replacement. For this integration to work, it is recommended that open-source platforms be used for compatibility. As an example, existing EHRs would be a great addition to AI systems adopted in pediatric health [36]. Some widely adopted open-source health information systems include District Health Information Software 2 for health management information, Open Medical Record System for electronic medical records, and Bahmni for hospital management. These platforms provide foundational platforms that can integrate AI applications while maintaining local adaptability. Such open-source platforms can facilitate collaborative development and promote application programming interface interoperability across diverse geographical contexts.

Privacy, Security, and Equity

Data privacy and security are fundamental to ethical health informatics and AI deployment in pediatric care. Some legal protections relating to the protection of personal information include the General Data Protection Regulation in Europe and the Health Insurance Portability and Accountability Act in the United States. However, many countries have distinct legal traditions and frameworks rooted in international human rights principles. Some of these include the right to privacy established in post–World War II conventions, such as the Universal Declaration of Human Rights and the Nuremberg Code's protections for research participants (Universal Declaration of Human Rights/Nuremberg Code). Such regulations and policies provide an important legal foundation for protecting personal information [37]. Technical measures that can be implemented include encryption for data at rest and in transit, role-based access controls, and audit trails. Privacy-enhancing technologies, such as federated learning and anonymization, enable collaborative model development without violating the patient's privacy [30]. Privacy protections must be complemented by informed consent processes that respect individual and community autonomy, including the right to opt out of data use for AI development. Health care organizations should set up independent ethics committees to monitor data protection and compliance with ethics rules.

Bridging the digital divide is critical to ensuring equitable access to health informatics and AI-driven pediatric care. Strategic investments in digital infrastructure, such as connectivity, computing resources, and technical capacity, are essential for wider adoption in underserved regions [24]. Mobile health platforms integrated with AI applications may offer scalable options for delivering health care services to areas with limited resource access. These platforms can support diagnostic and consultation services, potentially reducing the need for patients to travel to higher-level health care facilities [15]. Co-design approaches that engage local communities and health workers will bolster cultural appropriateness and language accessibility in pediatric practice.

Cost-Effective Implementation Strategies

Resource constraints present barriers to AI adoption in many settings. Hiring specialized experts during the initial phase of implementation of health informatics and AI, and the cost of training employees to acquire the new skills required for performing artificially intelligent tasks requires substantial resources [32]. Public-private partnerships may facilitate implementation in some contexts, but careful attention to governance, sustainability, and public interest protections is essential. Adopting open-source platforms may help reduce licensing costs and support transparent, collaborative development of AI tools, particularly in resource-constrained settings [30]. Another important measure is prioritizing AI applications to address high-burden conditions. For example, AI applications used in neonatal sepsis, pneumonia, and malnutrition would ensure resources are dedicated to areas with the greatest health impact.

XSL•FO
RenderX

Conclusions

This viewpoint has examined the intersection of health informatics and AI in pediatric health care. Using existing case studies and literature, we identified the current applications, benefits, and challenges of using health informatics and AI in pediatric health care. These technologies show promise in enhancing diagnostic accuracy, expanding access to care, and supporting more personalized treatment approaches in pediatric settings.

Significant implementation barriers persist, including challenges related to data quality, technical implementation, ethical governance, and persistent socioeconomic disparities that disproportionately affect LMICs. Addressing these challenges through collaborative, equity-driven efforts is essential to realizing AI's potential in enhancing global pediatric care. The tools exist, the need is urgent, and the time for global action is now.

Future research should explore long-term outcomes of AI implementations in diverse pediatric settings, develop context-appropriate validation frameworks, and establish governance structures that ensure ethical and equitable deployment.

Acknowledgments

The authors thank Victor Munyasya for providing helpful editorial feedback on early drafts of this manuscript. Victor is not affiliated with the Self Research Institute and was not involved in evidence selection or synthesis for this viewpoint. No generative artificial intelligence (AI) tools were used for the design, methodology, analysis, or writing of this manuscript. However, generative AI was used only to generate the table of contents image that accompanied the manuscript. All information for this use was provided during submission.

Funding

The authors declared that no financial support was received for this work.

Authors' Contributions

RSS and JM conceptualized the research. JM and RA helped structure and review the viewpoint. AM and GT provided helpful suggestions on the content and formatting of the paper. MM-K structured, reviewed, and formatted the viewpoint.

Conflicts of Interest

None declared.

References

1. Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019 Jul;8(7):2328-2331. [doi: 10.4103/jfmpc.jfmpc_440_19] [Medline: 31463251]
2. Gunasekaran V, Chamarthi VS. Artificial intelligence in pediatric clinical medicine: applications, innovations and practical implications. EJ-Clinicmed 2025;6(5):11-17. [doi: 10.24018/clinicmed.2025.6.5.398]
3. Ramgopal S, Horvat CM, Yanamala N, Alpern ER. Machine learning to predict serious bacterial infections in young febrile infants. Pediatrics 2020 Sep;146(3):e20194096. [doi: 10.1542/peds.2019-4096] [Medline: 32855349]
4. Coghlan S, Gyngell C, Vears DF. Ethics of artificial intelligence in prenatal and pediatric genomic medicine. J Community Genet 2024 Feb;15(1):13-24. [doi: 10.1007/s12687-023-00678-4] [Medline: 37796364]
5. Child mortality (under 5 years). World Health Organization. 2022 Jan 28. URL: https://www.who.int/news-room/fact-sheets/detail/child-mortality-under-5-years [accessed 2025-12-23]
6. Chen KC, Yu HR, Chen WS, et al. Diagnosis of common pulmonary diseases in children by X-ray images and deep learning. Sci Rep 2020 Oct 15;10(1):17374. [doi: 10.1038/s41598-020-73831-5] [Medline: 33060702]
7. Zhou M, Wu K, Yu L, et al. Development and evaluation of a leukemia diagnosis system using deep learning in real clinical scenarios. Front Pediatr 2021;9:693676. [doi: 10.3389/fped.2021.693676] [Medline: 34249819]
8. Padash S, Mohebbian MR, Adams SJ, Henderson RDE, Babyn P. Pediatric chest radiograph interpretation: how far has artificial intelligence come? A systematic literature review. Pediatr Radiol 2022 Jul;52(8):1568-1580. [doi: 10.1007/s00247-022-05368-w] [Medline: 35460035]
9. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol 2018 Jul 1;136(7):803-810. [doi: 10.1001/jamaophthalmol.2018.1934] [Medline: 29801159]
10. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. kBot: knowledge-enabled personalized chatbot for asthma self-management. Proc Int Conf Smart Comput SMARTCOMP 2019 Jun;2019:138-143. [doi: 10.1109/smartcomp.2019.00043] [Medline: 32832938]

XSL•FO
RenderX

11. Williams JB, Ghosh D, Wetzel RC. Applying machine learning to pediatric critical care data. Pediatr Crit Care Med 2018 Jul;19(7):599-608. [doi: 10.1097/PCC.0000000000001567] [Medline: 29727354]
12. Hashem H, Sultan I. Revolutionizing precision oncology: the role of artificial intelligence in personalized pediatric cancer care. Front Med (Lausanne) 2025;12:1555893. [doi: 10.3389/fmed.2025.1555893] [Medline: 40458648]
13. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
14. Guidance note: essential elements of quality in childcare settings (English). : World Bank Group; 2025 URL: https://documents1.worldbank.org/curated/en/099939304102517364/pdf/IDU-07686774-49fb-4c48-8fe2-91d966cf59ae.pdf [accessed 2026-04-08]
15. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018;3(4):e000798. [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]
16. Gaffar S, Gearhart AS, Chang AC. The next frontier in pediatric cardiology: artificial intelligence. Pediatr Clin North Am 2020 Oct;67(5):995-1009. [doi: 10.1016/j.pcl.2020.06.010] [Medline: 32888694]
17. Goel VV, Poole SF, Longhurst CA, et al. Safety analysis of proposed data-driven physiologic alarm parameters for hospitalized children. J Hosp Med 2016 Dec;11(12):817-823. [doi: 10.1002/jhm.2635] [Medline: 27411896]
18. Verder H, Heiring C, Ramanathan R, et al. Bronchopulmonary dysplasia predicted at birth by artificial intelligence. Acta Paediatr 2021 Feb;110(2):503-509. [doi: 10.1111/apa.15438] [Medline: 32569404]
19. Holmes JF, Yen K, Ugalde IT, et al. PECARN prediction rules for CT imaging of children presenting to the emergency department with blunt abdominal or minor head trauma: a multicentre prospective validation study. Lancet Child Adolesc Health 2024 May;8(5):339-347. [doi: 10.1016/S2352-4642(24)00029-4] [Medline: 38609287]
20. Fleischmann C, Reichert F, Cassini A, et al. Global incidence and mortality of neonatal sepsis: a systematic review and meta-analysis. Arch Dis Child 2021 Jul 19;106(8):745-752. [doi: 10.1136/archdischild-2020-320217] [Medline: 33483376]
21. Hodis B, Mughal S, Saadabadi A. Autism spectrum disorder. In: StatPearls: StatPearls Publishing; 2025. URL: https://www.ncbi.nlm.nih.gov/books/NBK525976 [accessed 2026-04-08] [Medline: 30247851]
22. Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. Transl Psychiatry 2012 Apr 10;2(4):e100-e100. [doi: 10.1038/tp.2012.10] [Medline: 22832900]
23. Sarkar MMH, Naser SR, Chowdhury SF, et al. M gene targeted qRT-PCR approach for SARS-CoV-2 virus detection. Sci Rep 2023;13(1):16659. [doi: 10.1038/s41598-023-43204-9]
24. Attallah O. MB-AI-His: histopathological diagnosis of pediatric medulloblastoma and its subtypes via AI. Diagnostics (Basel) 2021 Feb 20;11(2):359. [doi: 10.3390/diagnostics11020359] [Medline: 33672752]
25. Hravnak M, Chen L, Dubrawski A, Bose E, Clermont G, Pinsky MR. Real alerts and artifact classification in archived multi-signal vital sign monitoring data: implications for mining big data. J Clin Monit Comput 2016 Dec;30(6):875-888. [doi: 10.1007/s10877-015-9788-2] [Medline: 26438655]
26. Leite P, Chamba A, Zaarour Y, Leska-See D. A slippery slope: the opportunities and risks of digital approaches and technology in social protection systems. : World Bank Group; 2025 URL: https://hdl.handle.net/10986/42948 [accessed 2026-04-08] [doi: 10.1596/42948]
27. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020 May 16;395(10236):1579-1586. [doi: 10.1016/S0140-6736(20)30226-9] [Medline: 32416782]
28. Michelson KN, Klugman CM, Kho AN, Gerke S. Ethical considerations related to using machine learning-based prediction of mortality in the pediatric intensive care unit. J Pediatr 2022 Aug;247:125-128. [doi: 10.1016/j.jpeds.2021.12.069] [Medline: 35038439]
29. Castro HE, Kumar R, Malpica-Llanos T, et al. Sharing knowledge on innovative medicines for noncommunicable diseases: a compendium of good practices for sustainable access. : World Bank Group; 2018 Nov 21 URL: https://thedocs.worldbank.org/en/doc/160481542759870061-0090022018/render/MSHRTIGLOHICompendiumDraftNov20.pdf [accessed 2025-09-26]
30. Naik N, Hameed BMZ, Shetty DK, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? Front Surg 2022;9:862322. [doi: 10.3389/fsurg.2022.862322] [Medline: 35360424]
31. WHO and HL7 collaborate to support adoption of open interoperability standards. World Health Organization. 2023 Jul 3. URL: https://www.who.int/news/item/03-07-2023-who-and-hl7-collaborate-to-support-adoption-of-open-interoperability-standards [accessed 2025-10-19]
32. Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. NPJ Digit Med 2022 Oct 28;5(1):162. [doi: 10.1038/s41746-022-00700-y] [Medline: 36307479]
33. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]

XSL•FO
RenderX

34. Jamil Abusamra HN, Ali SHM, Khidir Elhussien WA, Ahmed Mirghani AM, Alameen Ahmed AA, Abdelrahman Ibrahim ME. Ethical and practical considerations of artificial intelligence in pediatric medicine: a systematic review. Cureus 2025 Feb;17(2):e79024. [doi: 10.7759/cureus.79024] [Medline: 40104456]
35. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. World Health Organization. 2025 Mar 25. URL: https://www.who.int/publications/i/item/9789240084759 [accessed 2026-04-28]
36. Paton C, Braa J, Muhire A, et al. Open source digital health software for resilient, accessible and equitable healthcare systems. Yearb Med Inform 2022 Aug;31(1):67-73. [doi: 10.1055/s-0042-1742508] [Medline: 35654431]
37. Pistilli G, Trevelin B. Can AI be consentful? arXiv. Preprint posted online on Jun 27, 2025. [doi: 10.48550/arXiv.2507.01051]

Abbreviations

ADOS: Autism Diagnostic Observation Schedule
AI: artificial intelligence
ASD: autism spectrum disorder
BPD: bronchopulmonary dysplasia
CT: computed tomography
EHR: electronic health record
LMICs: low- and middle-income countries
ML: machine learning
NLP: natural language processing
PECARN: Pediatric Emergency Care Applied Research Network
WHO: World Health Organization

Edited by K Gierend; submitted 09.Aug.2025; peer-reviewed by D Walker, S Bitaraf; revised version received 16.Feb.2026; accepted 07.Mar.2026; published 13.May.2026.

Please cite as:
Mbogori-Kairichi M, Snead RS, Marczika J, Ambalavanan R, Towett G, Malioukis A
Transforming Pediatric Care Through AI: Bridging the Digital Divide in Health Informatics
Online J Public Health Inform 2026;18:e82108
URL: https://ojphi.jmir.org/2026/1/e82108
doi:10.2196/82108

© Mercy Mbogori-Kairichi, R Sterling Snead, Julia Marczika, Radha Ambalavanan, Gideon Towett, Alex Malioukis. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 13.May.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Trend in Devices and Digital Tools for Remote Consultation From Medical Providers to Specialists: Scoping Review

Risa Hara[1], MD; Yosuke Hirakawa[1], MD, PhD; Yuka Sugawara[1], MD, PhD; Masao Iwagami[2], MD, PhD; Ryota Inokuchi[3], MD, PhD; Masaomi Nangaku[1], MD, PhD

[1]Division of Nephrology and Endocrinology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
[2]Department of Digital Health, Social Medicine Group, Institute of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan
[3]Department of Clinical Engneering, The University of Tokyo Hospital, Bunkyo, Tokyo, Japan

Corresponding Author:
Yosuke Hirakawa, MD, PhD
Division of Nephrology and Endocrinology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Background: Digital health care technologies, including mobile applications and telemedicine platforms, have transformed how medical professionals communicate and deliver care. Remote consultation by doctors plays a vital role in ensuring access to appropriate expertise, particularly in medically underserved or geographically remote areas. However, the diversity in technological modalities, devices, and patterns of use across specialties and regions has not been systematically mapped.

Objective: This study aimed to explore the current status and characteristics of teleconsultations among medical providers and specialists, focusing on device use, consultation modalities, clinical specialties, and regional differences. Through this approach, we aimed to provide a comprehensive overview of technological and practical trends in mobile health (mHealth) and telemedicine.

Methods: A systematic scoping review was conducted in accordance with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines using the MEDLINE and Embase databases. The search covered studies published up to February 2026, with no restrictions on the publication year. Studies meeting the predefined inclusion criteria were also included.

Results: A total of 255 citations were screened, and 99 articles were included. Studies were analyzed according to consultation method, target, and regional characteristics. Of these, 81 out of 99 (81.8%) articles were in the doctor-to-doctor category. Email, web, or app-based platforms were the most common. In addition, 86 out of 99 studies used medical images, most frequently photographs. Orthopedics and dermatology were the most frequently involved specialties, followed by internal medicine. Regarding the region, 60 of 99 studies were domestic, and the other 39 studies were international. Rural-to-urban domestic consultations comprised 27 out of 99 (27.3%) studies, whereas consultations from low- and middle-income countries and high-income countries accounted for 11 of 99 (11%) studies.

Conclusions: This review examined doctor-to-doctor and doctor-to-patient consultations with doctor involvement. Specialties in which medical images are central, such as orthopedics and dermatology, were more frequently represented than in other fields. This highlights disparities in the use of teleconsultation across clinical disciplines and suggests that addressing these imbalances is essential for broader adoption. Furthermore, the findings indicated a progressive shift from videoconference-based interactions to mobile and app-based platforms, reflecting ongoing technological advancements. Optimizing the integration of these digital tools and promoting equitable access are critical for enhancing the quality and reach of teleconsultation practices in future digital health systems.

(Online J Public Health Inform 2026;18:e87559) doi:10.2196/87559

KEYWORDS

telemedicine; teleconsultation; remote consultation; medical consultation; doctor-to-doctor consultation; doctor-to-patient with doctor consultation; digital technology; mobile health

Introduction

Background

Teleconsultation is a form of telemedicine in which specialists and other health care professionals provide advice and clinical support to physicians and patients using information and communication technologies [1]. Such consultations play a crucial role in ensuring appropriate care, particularly when specialty departments are unavailable, or services are delivered in underserved or remote areas. Teleconsultation applies to

XSL•FO
RenderX

diverse scenarios including routine outpatient care, emergency response, and multiple specialties. Their importance has grown since the COVID-19 pandemic, accelerating remote medical care and creating new health care delivery models [2]. Teleconsultation has demonstrated strong potential for bridging gaps in access where specialists are scarce.

Teleconsultation is commonly divided into two types: doctor-to-doctor consultations, in which physicians seek guidance from specialists, and doctor-to-patient consultations with doctor involvement, in which a physician mediates communication between the patient and specialist [3]. Implementation methods include telephone calls, email, and web or app-based platforms including store-and-forward or synchronous styles. Devices and platforms continue to evolve with technology, and the choice often depends on infrastructure, privacy requirements, communication environment, and clinical context [4].

In recent years, the rapid advancement in digital health technologies has diversified the implementation environment for telemedicine. These innovations have enabled rapid and detailed teleconsultation across geographic and institutional boundaries, expanding access to specialist expertise and transforming the conduct of remote collaboration in health care.

Against this backdrop, this study conducted a scoping review of the current status of teleconsultation to clarify the technological and practical trends in the evolving landscape of digital and remote medical communication.

Objectives

This study aimed to clarify the technological trends and clinical characteristics in doctor-to-doctor and doctor-to-patient consultations involving doctors. We conducted a scoping review of doctor-to-doctor and doctor-to-patient consultations with doctor involvement. We comprehensively analyzed the implementation methods, devices used, and clinical scenarios, in order to clarify the challenges and opportunities for expanding the use of teleconsultation in the future.

Methods

Study Design

This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, and the detailed methodology is provided in Checklist 1.

Data Sources and Search Strategy

A systematic search was performed using the MEDLINE and Embase databases. The search covered studies published up to February 28, 2026, with no restrictions on the publication year. The search strategy was designed to capture studies reporting teleconsultations from doctors to doctors. Search terms included: MEDLINE: (telemedicine OR “online medical care” OR “teleconsultation” OR “online consultation” OR “telemedical consultation”) AND (“D to D” OR “doctor to doctor” OR “physician to physician”); Embase: (telemedicine OR “online medical care” OR teleconsultation OR “online consultation” OR “telemedical consultation”) AND (“D to D” OR “doctor to doctor” OR “physician to physician”). Details of the search strings are provided in Multimedia Appendix 1. Gray literature was excluded from this review. In addition, a single backward citation search was performed on all identified review articles to broaden coverage. We also conducted a robustness check using only the results of the database search to confirm that the findings were consistent with those of the overall analysis.

Study Selection

After deduplication, titles and abstracts were screened using Covidence software (Veritas Health Innovation) by 2 independent reviewers (RH and YH). Full-text articles were then assessed using predefined inclusion and exclusion criteria. Studies were included if they reported teleconsultation between medical providers and specialists, regardless of whether patients were present. The exclusion criteria were as follows: (1) protocols and proposals, (2) systematic reviews, commentaries, and scoping reviews, (3) non-full-text publications, (4) publications published after February 2026, and (5) articles that were not published in English. Disagreements were resolved by consensus with a third reviewer (YS). The Cohen kappa coefficient was calculated to ensure the reliability of the study selection process. In this assessment, the kappa value was calculated using the following formula: kappa = (Po – Pe) / (1 – Pe), where Po represents the observed proportion of agreement, and Pe represents the hypothetical probability of chance agreement between the 2 independent reviewers (YH and RH). The results indicated substantial agreement between these 2 reviewers (YH and RH), with a kappa value of 0.603 for the full-text reviews. These results indicate substantial agreement between reviewers (YH and RH), suggesting that the screening process was conducted with a high degree of reliability. Discrepancies were resolved by a third reviewer.

Data Extraction

Data extraction was conducted independently by two reviewers (RH and YH) using Covidence software (Veritas Health Innovation) and Excel (Microsoft Corp). Extracted data included: (1) citation details, (2) country of origin, (3) geographic characteristics (domestic or international, urban or rural), (4) medical specialties, (5) patient characteristics (adult or child), (6) evaluation criteria, (7) consultation methods (eg, email and telephone), (8) use of medical images, (9) inclusion of patients in consultations, and (10) reported costs. This review has been registered in the Open Science Framework Registry [5].

Terminology and Classification Criteria

Due to variations in definitions, this study classified consultations as rural only when the original articles explicitly described the setting as geographically remote or lacking medical resources. Consultations conducted in remote areas were not considered rural unless these characteristics were clearly stated. Low- and middle-income countries (LMICs) and high-income countries (HICs) were classified according to World Bank classification. We categorized consultation methods by using web or app-based platforms into “Web or app-based synchronous platforms” for real-time engagement and “Web or

XSL•FO
RenderX

app-based store-and-forward platforms" for asynchronous data transmission via integrated chat or email-based interfaces, because consultation modalities should be categorized into synchronous (real-time) and asynchronous (store-and-forward) interactions. In addition, when consultations were conducted synchronously via modalities such as videoconferencing, we changed the designation from "Store-and-forward video" (asynchronous) to "Real-time video" (synchronous) and performed the analysis again.

Data Analysis

Fisher exact test was performed to evaluate the differences between domestic and international teleconsultations both in Table 1 and Table S2 in (Multimedia Appendix 2). A *P* value of ≤.05 was considered statistically significant. No adjustments were made for multiple comparisons; therefore, all statistical analyses were exploratory.

Table . Methods of consultation by domestic or international consultation. Some studies used multiple devices or modalities; therefore, the total number of modality occurrences exceeded that of the number of included studies (n=99) because some studies used multiple modalities. Counts reflect individual modality occurrences rather than unique studies.

Characteristics	Domestic consultations (n=60)	International consultations (n=39)	*P* value
Web or app-based synchronous platforms	25 (40.3)	15 (34.9)	.69
Email	15 (24.2)	19 (44.2)	.04[a]
Web or app-based store-and-forward platforms	11 (17.7)	6 (14)	.78
Phone	6 (9.7)	1 (2.3)	.24
Other	4 (6.5)	1 (2.3)	.65
Unknown	1 (1.6)	1 (2.3)	≥.99

[a]*P*<.05.

Review Period

The review process was conducted between March 2024 and March 2026.

Ethical Considerations

This study is a scoping review of previously published literature. Ethical approval was not required because no human participants or identifiable personal data were involved.

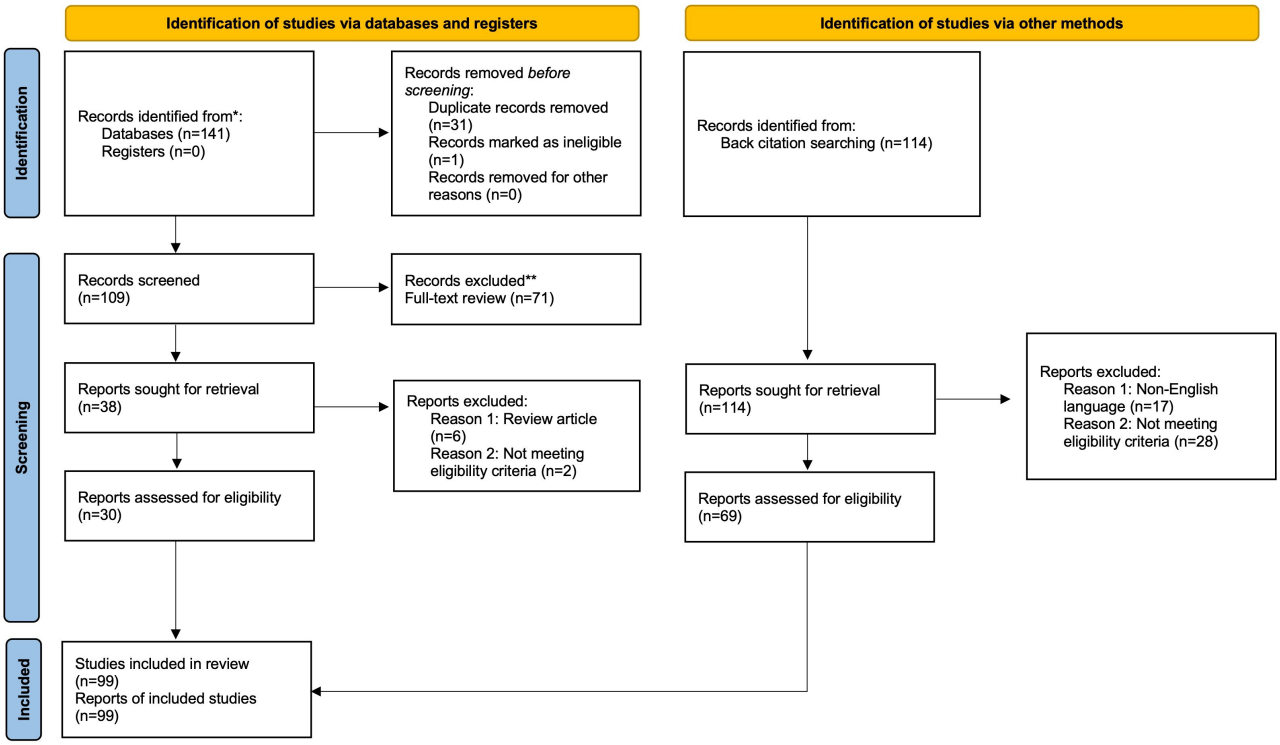
Results

Literature Search

Initially, 141 articles were identified through database searches. After deduplication, 109 records were selected for screening. The screening of title and abstract reduced the number of potentially eligible articles to 38. Among these, 6 review articles and 2 articles were excluded after a full-text review, leaving 30 studies. A backward citation search of the 6 review articles identified during the full-text review yielded an additional 114 articles for full-text review. Among these, 28 were excluded based on the eligibility criteria, and 17 were excluded for publications in non-English languages. Finally, a further 69 articles were included. Overall, 99 studies [6-104] (Multimedia Appendix 3) were included in the qualitative analysis. A PRISMA-ScR flow diagram summarizing the study selection process is shown in Figure 1. A list of all 99 studies and their information is provided in Multimedia Appendix 3. After excluding backward citation studies, the total number of included studies was 30 [9-18,20-25,86-97,104].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for new systematic reviews which included searches of databases, registers, and other sources.



Evaluation Items of the Included Studies

Following the study selection process, we extracted and categorized key entities from each included article based on the following criteria: (1) citation details (authors and year of publication); (2) country of origin; (3) geographic characteristics, including whether the study was domestic or international and whether it was conducted in urban or rural settings; (4) target medical departments or organs; (5) characteristics of the target population, including sample size and whether participants were adults or children; (6) evaluation criteria used in each study (such as clinical outcomes, user satisfaction, or feasibility); (7) consultation method, including email, telephone call, videoconferencing, or app-based platforms; (8) whether medical images were used, and if so, what types of images were used (eg, photographs and radiological images); (9) whether the consultation was conducted exclusively among health care professionals or whether patients were also involved; and (10) cost reporting and financial aspects, if these were referenced in the study.

In cases in which specific data were unavailable in the original publication, the information was reported as missing. These extracted data points formed the basis for the categorization and analysis presented in the subsequent sections.

Study Description

The characteristics of the studies are shown in Table 2. All 99 articles were published between 1996 and 2026. In particular, the COVID-19 pandemic in 2020 accelerated the global adoption of medical practices designed to minimize face-to-face interactions through telemedicine. The countries in which the articles were published are indicated by region in Table 2. North America accounted for the largest number of articles (30/99, 30.3% studies 9,12,15,18,21,23,25,27,31,33,35,47,49-53,55-57,69,74-79,93,101,103]), followed by Europe (17 studies [10,20,22,29,42,44-46,48,54,58,62,65,70,72,85,87], 17.2%). Overall, 19 (19.2%) studies [14,19,39-41,43,60,63,64,71,73,81-84,91,92,97,104] reported on cross-national or international consultations. Owing to the nature of the telemedicine field, 3 (3%) studies [7,67,68] were conducted in Antarctica, which lacks medical facilities and staff. The analysis, excluding studies identified through backward citations, was generally consistent with the main results (Table S1 in Multimedia Appendix 2).

XSL•FO
RenderX

Table . Study characteristics. Percentages are rounded to one decimal place; component sums may not equal the total. The total number of modality occurrences and target medical departments exceeded the number of included studies (n=99) because some studies included multiple consultation characteristics. Counts reflect individual modality occurrences rather than unique studies.

Characteristic	Studies, n (%)
Publication year	
1996 - 2000	3 (3)
2001 - 2005	5 (5.1)
2006 - 2010	15 (15.2)
2011 - 2015	34 (34.3)
2016 - 2020	17 (17.2)
2021 - 2025	24 (24.2)
2026	1 (1)
Geographic area of teleconsultation activity	
North America	30 (30.3)
South America	2 (2)
Asia	15 (15.2)
Europe	17 (17.2)
Middle East	4 (4)
Africa	4 (4)
Oceania	5 (5.1)
Antarctica	3 (3)
International	19 (19.2)
Target medical department	
Orthopedics	22 (21.2)
Dermatology	19 (18.3)
Internal medicine	15 (14.4)
Neurology	5 (4.8)
Cardiology	3 (2.9)
Infectiology	3 (2.9)
Oncology	1 (1)
Pulmonology	1 (1)
Hepatology	1 (1)
Pain medicine	1 (1)
Surgery	7 (6.7)
Emergency room	6 (5.8)
Ophthalmology	6 (5.8)
Pediatrics	4 (3.8)
Obstetrics	1 (1)
Oral surgery	1 (1)
Psychiatry	1 (1)
Pathology	1 (1)
Radiology	1 (1)
Not specified or multidisciplinary	20 (19.2)
Consultation direction	
Primary care provider to specialist	62 (62.6)

XSL•FO
RenderX

Characteristic	Studies, n (%)
Specialist to specialist	36 (36.4)
Both of the above	1 (1)
Patient involvement	
Doctor to doctor	81 (81.8)
Doctor to patient consultation with doctor	14 (14.1)
Both of the above	4 (4)
Domestic or international consultations	
Domestic consultations	60 (60.6)
Rural to urban area	27 (27.3)
Not specified	33 (33.3)
International consultations	39 (39.4)
LMICs[a] to HICs[b] area	11 (11.1)
Military installations	10 (10.1)
Antarctica	3 (3)
Not specified	15 (15.2)

[a]LMICs: low- and middle-income countries

[b]HICs: high-income countries.

Domestic or International Consultation

Of the 99 included articles, 60 (60.6%) studies [6,8,10-13,15-18,20-27,29,30,32-36,39,41-58,62,85-88,93-96,98-103] reported domestic consultations, and the remaining 39 (39.4%) studies [7,14,19,28,31,37,38,40,59-61,63-84,89-92,97,104] reported international consultations (Table 2). Among domestic cases, 27 studies [11,12,14-18,21,26,28-33,37,55,58,62,85-87,92,94-96] described communication from rural to urban areas, particularly consultations from rural clinics to urban hospitals. Consultations from LMICs to HICs accounted for 11 of the international cases, reflecting the role of teleconsultation in addressing global disparities in access to specialist care.

Target Medical Departments

As summarized in Table 2, the included studies covered a wide range of medical specialties. Among the 99 studies, the most frequently targeted departments were orthopedics (22, 21.2% studies [29-46,54,79,97,100]) and dermatology (19, 18.3% studies [12,15,17,25-28,47-53,66,82,83,85,102]). Internal medicine was the third most frequently reported field (15, 14.4% studies [6,7,9,13,22-24,58,60,62,63,73,77,78,97]). Within this field, neurology was the most common subspecialty, followed by cardiology (3, 2.9% studies [7,62,77]) and infectious diseases (3, 2.9% studies [13,73,78]). Neurology consultations are often concerned with emergency transport cases such as cerebral hemorrhage and stroke [6], and specialist consultations for conditions such as headache and tremor [105]. In contrast, cardiology consultations are frequently conducted for urgent conditions, including pericardial effusion secondary to pericarditis, often using tele-echocardiographic evaluation [7], and commonly involve teleconsultations based on readily shareable diagnostic data, such as electrocardiograms and chest radiographs [8].

In addition, consultations were reported in surgery (7, 6.7% studies [54,55,59,61,64,70,97]), ophthalmology (6, 5.8% studies [11,20,76,81,93,96]), pediatrics (4, 3.8% studies [38,65,87,101]), including pediatric surgery, pediatric orthopedics, and pediatric oncology, and emergency medicine (6, 5.8% studies [18,56,57,71,74,98]). One article each reported on pathological and psychiatric consultations. The remaining 20 studies [8,10,14,16,19,67-69,72,75,84,88-92,95,99,103,104] primarily addressed the technical aspects of teleconsultation or discussed teleconsultation without specifying the disease area.

Consultation Direction and Patient Involvement

The directional flow and structural forms of teleconsultation varied across the included studies. We categorized the consultations into two primary types based on direction: Of all 99 studies, primary care provider-to-specialist and specialist-to-specialist. Most studies (62/99, 62.6% studies [6-19,21,23-27,29-31,33,34,37,40,41,47-51,53-55,57,61,67-69,72,74-79,84,85,87,88,92,93,95-99,101,103,104]) described primary care provider–to–specialist consultations in which general practitioners or community-based physicians sought expert input for diagnostic clarification or treatment decisions. These interactions often support primary care providers in underserved rural regions.

In contrast, 36 of 99 (36.4%) studies [10,20,22,28,32,35,36,38,39,42-46,52,56,58-60,62-66,70,71,73,80-83,86,88,94,100,102] reported specialist-to-specialist consultations, in which subspecialists communicated with other specialists; for example, a general surgeon consulted a radiologist or dermatologist. One (1%) study [10] described both consultation styles.

Regarding the structural form of consultations, we also evaluated for direct patient involvement. A substantial proportion of the

XSL•FO
RenderX

studies (81/99, 81.8% studies [6,9,10,12-23,25-31,33,35,36,38-42,47-55,57-70,72-79,81-85,89-92,94-104]) involved doctor-to-doctor interactions conducted exclusively among health care professionals without direct patient participation. These models typically entail the exchange of clinical data, imaging data, or summary reports via electronic platforms.

Meanwhile, 14 of 99 (14.1%) studies [7,8,11,24,32,34,37,43-46,56,71,80] featured doctor-to-patient with-doctor models, in which a physician facilitated communication between the patient and a remote specialist. Such arrangements are often used in geographically isolated or low-resource settings where direct specialist access is unavailable.

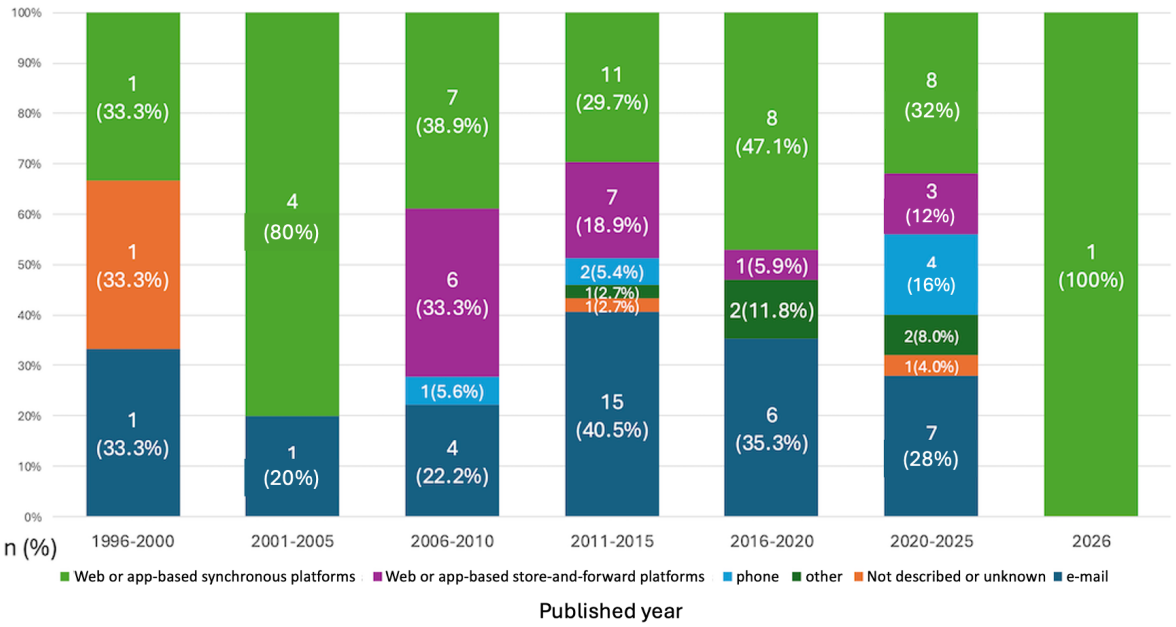
When comparing doctor-to-doctor and doctor-to-patient consultations with doctor involvement, domestic cases were predominantly conducted as doctor-to-doctor consultations, whereas international cases were markedly more frequent in the doctor-to-patient with doctor format. North America showed a significantly higher proportion of doctor-to-doctor consultations, whereas Oceania had significantly more doctor-to-patient consultations involving doctors. Videoconferencing was significantly more frequent in doctor-to-patient consultations with doctor involvement, whereas email, although not statistically significant, tended to be more common in doctor–to–doctor consultations. In terms of imaging modalities, photographs were significantly more frequent in doctor-to-doctor consultations, whereas videos were significantly more common in doctor-to-patient consultations with doctor involvement. No significant differences were observed for other imaging modalities; however, no cases in the doctor-to-patient with doctor group lacked imaging usage. Among the clinical specialties, orthopedics was significantly more frequent in doctor-to-patient consultations with doctor involvement, whereas dermatology was significantly more common in doctor-to-doctor consultations, with no significant differences observed for other specialties.

Methods of Consultation

All the articles were classified according to the methods used for teleconsultation, including email, telephone, web or app-based synchronous platform, web or app-based store-and-forward platform, other methods, and articles in which the method could not be identified.

Figure 2 presents a color-coded vertical bar chart of the consultation methods for the five-year period from 1996 to 2026. Figure S2 in Multimedia Appendix 4 also presents the same bar chart in 30 articles, excluding backward-searching studies. Email was used consistently since 2001, and a significant shift toward web or app-based synchronous platforms was observed in more recent years. Web or app-based platforms include independently developed web services within specific countries or hospital groups, reflecting the broader global adoption of such digital interfaces.

Figure 2. Time trends of in consultation methods; consultation methods across five-year periods from 1996 to 2026.



The differences in teleconsultation methods between domestic and international consultations are presented in Table 1 and Table S2 (Multimedia Appendix 2) with and without backward citation studies, respectively. Email was significantly more common in international consultations than in domestic consultations (P=.04), while when backward citation studies were excluded, the difference was not statistically significant. Reflecting the unit of analysis as modality occurrences, the total count exceeded the number of included studies because some programs used multiple methods.

XSL•FO
RenderX

Image Modalities

Of the 99 studies reviewed, 86 (86.9%) [6-9,11,12,15-18,20,24-73,75-88,90-94,96-98,100-102] reported the use of some form of medical imaging during consultations. Among these 86 studies, clinical photographs (n=53, 31.9% [11,12,15,17,18,25-28,36-38,40,41,47-53,59,61,63,64,66,67,69-73,75-79,81-85,88,90,92-94,96,98,100-102]), real-time video (n=36, 21.7% [6-8,11,24,29-31,34-37,39,43-46,54-60,62,64,65,69,71,80,86-88,90-92]), and radiographic imaging (n=44, 36.5% [6,7,9,16,29,30,32-34,36-38,40-46,54,59,61,62,72,75,79-81,96,97,100]) were the most common modalities, followed by magnetic resonance imaging, specialized imaging, ultrasonography, physiological data, store-and-forward video, pictures, and pathological imaging. As some studies used multiple modalities, the total number of identified modality occurrences was 166. The details are summarized in Table 3. Radiographic imaging included radiography, computed tomography, and angiography, whereas specialized imaging included dermoscopic images in dermatology and eye fundus photographs in ophthalmology. Also, physiological data included electrocardiography, pulmonary function tests, and cardiotocography, and pictures included pictures drawn by patients and burn diagrams. Analyses excluding studies identified through backward citation yielded generally consistent results (Table S3 in Multimedia Appendix 2).

Table . Technical and implementation features. Some studies used multiple devices or modalities; therefore, the total number of modality occurrences exceeded that of the number of included studies (n=99) because some studies used multiple modalities. Counts reflect individual modality occurrences rather than unique studies.

Image modalities use	Studies, n (%)
Yes	86 (86.9)
Photograph	53 (31.9)
Radiographic imaging	44 (26.5)
Real-time video	36 (21.7)
Specialized imaging	8 (4.8)
Ultrasound imaging	7 (4.2)
MRI[a]	6 (3.6)
Physiological data	5 (3)
Pathological imaging	3 (1.8)
Picture	3 (1.8)
Store-and-forward video	1 (0.6)
Unknown, no description provided	9 (9.1)
None	4 (4)

[a]MRI: magnetic resonance imaging.

The most frequently used modality, clinical photography, was used in 53 studies [11,12,15,17,18,25-28,36-38,40,41,47-53,59,61,63,64,66,67,69-73,75-79,81-85,88,90,92-94,96,98,100-102], particularly in dermatology and orthopedic surgery. Photography allows for rapid and clear communication of visible physical symptoms such as rashes, wounds, joint deformities, and skin lesions. With the advent of high-resolution cameras and secure image-sharing applications, teleconsultation has been increasingly conducted using high-quality images in a simplified and privacy-conscious manner. The remaining 4 of 99 (4%) studies [13,14,21,99] did not use any medical images during their consultation process, and 9 (9.1%) studies [10,19,22,23,74,89,95,103,104] provided no description of the modalities used.

Discussion

Principal Results

This scoping review synthesized 99 studies (published until February 2026) to provide an overview of current teleconsultation practices across clinical settings, regions, and modalities. We observed substantial variations in implementation according to specialty, consultation method, and geographical context. Within the context of doctor-to-doctor consultations, this pattern suggests that image-intensive specialties may be more compatible with remote consultation formats; therefore, these models were adopted more readily.

Technological Evolution and Communication Modalities

Our review identified a gradual shift from telephone to email and web or app-based platforms (both store-and-forward or synchronous type), reflecting broader advances in the telemedicine infrastructure [106]. These platforms enable seamless transmission of images and clinical data, thereby enhancing the quality of remote evaluation, and are especially common in international consultations, as their asynchronous format overcomes time-zone barriers.

Web or app-based platforms offering structured input and secure image sharing are well-suited to image-dependent specialties [107]. With the widespread use of mobile devices and growth of mobile health (mHealth) platforms, teleconsultations have

XSL•FO
RenderX

become increasingly accessible to health care professionals, even in resource-limited environments. These trends indicate that technology accessibility and platform functionality are central to teleconsultation.

Image Modalities in Teleconsultation

A notable finding of this review is the high frequency of medical imaging use in teleconsultations, particularly in clinical photographs and radiological imaging. These images are not only common but also integral to the consultation process. Among the included studies, the majority used some form of visual data to support diagnosis or decision-making. Photographs of skin lesions, wounds, or orthopedic conditions are frequently transmitted via email or app-based platforms, enabling remote specialists to provide accurate advice without direct patient contact. Radiological images, including x-ray and computed tomography scans, have been used in many consultations, particularly in orthopedics and emergency care.

Recently, the potential use of 3D imaging technology for teleconsultation and telementoring in oral and maxillofacial surgeries has been investigated [106]. Compared to traditional 2D images, 3D images and virtual-reality applications created from 3D data can enable clinicians to understand complex anatomical structures more clearly and share surgical plans more effectively. At present, creating 3D data requires expensive equipment and software, as well as a fast and stable Internet connection for data transfer. However, this technology has already been adopted to support communication and education among doctors in remote areas [108][109]. In addition, mobile-based image capturing and AI-supported image analysis tools have been increasingly integrated into teleconsultation workflows across multiple clinical specialties, including ophthalmology [110]. Collectively, these developments highlight the critical role of visual diagnostic tools in remote clinical settings. Improving the accessibility of image capturing and sharing, particularly in resource-limited settings, can expand teleconsultation to a broader range of clinical scenarios.

Clinical Specialties and Image-Driven Practices

We found a significant disparity in teleconsultation activities across clinical departments. Medical fields such as orthopedics and dermatology accounted for a substantial proportion of the included studies. These specialties frequently rely on visual cues, and clinical decisions are often made based on images alone, particularly clinical photographs or radiological images. This pattern suggests that image-intensive specialties are more compatible with remote consultation formats and have, therefore, adopted teleconsultation more readily. In a robustness check excluding studies identified through backward citations, no significant differences were observed across specialties (Table S1 in Multimedia Appendix 2), whereas the proportion of orthopedics was somewhat lower, suggesting a potential influence of citation-based inclusion. Fields such as internal medicine are relatively underrepresented compared with orthopedics and dermatology. This may be because these specialties often require more integrative assessments involving history-taking, laboratory data, and physical examinations, which are difficult to fully replicate in a remote setting. Several studies have suggested that the adoption of telemedicine varies across medical specialties and clinical conditions, largely due to differences in the need for physical examinations [111]. The increasing use of mobile communication tools and wearable sensors may help expand teleconsultation to nonimage-based specialties by enabling the real-time transmission of physiological or biometric data.

The Difference Between General Telemedicine and Teleconsultation

As general telemedicine and the types of teleconsultations examined in this study have different goals and evaluation criteria, the mix of specialties involved and reasons for adoption are not the same.

In the United States, since the COVID-19 pandemic, telemedicine has been used more frequently in history-focused fields such as psychiatry, endocrinology (particularly diabetes care), and neurology [112]. It has been widely used in pediatrics and gynecology. One reason is that childcare responsibilities and work demands often make clinic visits difficult. Telemedicine reduces travel and time burdens, thereby enabling more visits to be completed [113]. Telemedicine delivers care directly to patients in real time and is typically evaluated based on visit completion rates, reduced missed or interrupted visits, and patient satisfaction [114].

In contrast, teleconsultation aims to obtain specialist advice to improve diagnostic confidence, shorten patient wait times, support primary care decision-making, and, in some cases, avoid in-person referrals or transfers. Teleconsultation is particularly common in specialties that rely heavily on imaging, such as dermatology and orthopedics.

Opportunities and Remaining Challenges

Several challenges must be addressed to support the continued expansion of teleconsultation. First, underrepresented specialties require greater integration. Even in fields where image-based diagnostics are uncommon, transmitting laboratory results or PDF reports can help bridge this gap. Devices and platforms should be designed with flexibility and interoperability to accommodate diverse clinical data formats [4].

Recent studies in nephrology have illustrated both the obstacles and potential solutions for nonimage-intensive specialties. Our previous study [115] noted that telemedicine in nephrology is underused because of difficulties in conducting physical examinations, laboratory tests, and remote imaging. Promising approaches include estimating serum creatinine from saliva or tear fluid, measuring hemoglobin via smartphone photographs of the conjunctiva or fingernails, and using motion-capture–assisted kidney ultrasound outside hospital settings. These innovations can be adapted more broadly within the field of internal medicine to enhance remote diagnostics.

This review also identified 29 studies [8,9,12-14,18,19,30,33,34,41,43,44,47,54,59,69,77,79,82,87-89,92,93,95,96,101,104] referencing cost-related benefits, such as reduced patient transport and faster diagnostic decisions [9]. These findings highlight the economic advantages, although more systematic cost-effectiveness analyses are needed.

XSL•FO
RenderX

Finally, as teleconsultation expands globally, disparities in reimbursement models and regulations present barriers [116]. Although countries such as the United States have established broader billing systems, such as current procedural terminology (CPT) codes for interprofessional eConsults, others, including Japan, still face implementation challenges. Although Japan has integrated telemedicine into its Medical Care Act and has established specific medical fees for certain tele-diagnostic services; nonetheless, the reimbursement rates for most physician-to-physician consultations remain relatively low, and eligibility criteria are often too restrictive to support diverse interprofessional interactions across clinical specialties, which results in a more fragmented system than those with standardized, cross-specialty billing mechanisms. The development of coordinated international frameworks and standardized protocols is required for equitable and sustainable implementation [117].

Limitations

Although our scoping review was thorough, it had certain limitations that should be acknowledged. First, we searched the MEDLINE and Embase databases and included peer-reviewed full-text articles published in English. Therefore, studies published in other languages and the gray literature were not included, which may have reduced the comprehensiveness of the review. Specifically, this search strategy may have underrepresented teleconsultation initiatives in low-resource settings and international programs, where activities are more frequently documented in regional languages or unpublished institutional reports. Consequently, the global trends identified in this review should be interpreted as being primarily reflective of peer-reviewed, English-language literature rather than of a truly exhaustive global record. Second, a backward citation search was conducted to supplement the database search. This method may have introduced a selection bias by favoring studies that were more frequently cited or belonged to similar research networks, thereby affecting the overall balance of the study results. We conducted the same analyses, excluding studies identified through backward citation and confirmed that most of the evaluated characteristics showed no significant differences, whereas significant differences were observed in publication year ($P<.001$) and consultation direction ($P=.02$), owing to the retrieval of older literature and specific consultation types through backward searching (Tables S1-S3 in Multimedia Appendix 2). Third, narrative synthesis is a secondary analysis of data that focuses on the interpretations presented by the authors of original papers and is not based on primary data. Consequently, our findings represent an interpretation of the evidence and should be viewed as a heuristic framework, rather than as a definitive conclusion. Finally, this scoping review may reflect published research on teleconsultation rather than the full scope of real-world implementation because several telemedicine practices and the latest technological or regulatory shifts may not be formally documented in the literature or may be missed owing to the inherent time lag between practical innovation and academic publication. As such, the specialty distributions observed here could be influenced by research activity and publication patterns (including backward snowballing) as much as, or perhaps more than, the actual service uptake.

Conclusions

This scoping review highlights the diverse practices and evolving nature of teleconsultation, with particular emphasis on doctor-to-doctor and doctor-to-patient-with-doctor models. We found that image-based specialties such as orthopedics and dermatology are leading the implementation of teleconsultation, supported by the increasing use of image-compatible platforms such as email and mobile applications. The integration of visual data plays a crucial role in enabling remote clinical decision-making.

Despite these advancements, significant disparities remain across clinical fields and expanding teleconsultation into less image-dependent specialties will require innovations in data sharing and system design. Future progress in teleconsultation should focus on improving interoperability and integrating mobile technologies to promote equal access across regions and clinical areas. Continued innovation in digital health and supportive policies is essential for establishing teleconsultations as a stable and essential part of future health care systems.

Acknowledgments

The authors confirm that no generative AI or AI-assisted technologies were used in the writing of this manuscript, data analysis, or the preparation of any figures or tables.

The authors would like to thank Editage for English language editing.

Funding

This study was supported by the Ministry of Health, Labour and Welfare (Japan) (grant number 23IA2001), Grant-in-Aid for Scientific Research(C) (Japan) (grant number 25K11536), and the Japan Agency for Medical Research and Development (Japan) (grant number JP256f0137004).

Data Availability

All data analyzed during this study are included in the article and its appendices. The full list of the 99 included studies and their primary characteristics is provided in Multimedia Appendix 3.

XSL•FO
RenderX

Authors' Contributions

Study conception and design: YH, YS, MN
Investigation: RH, YH, YS, IM, MN
Writing the original draft: RH, YH
Writing the review and editing: YS, IM, MN
Funding acquisition: MN
Supervision: MN

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the search strings of this scoping review.
[PDF File, 16 KB - ojphi_v18i1e87559_app1.pdf]

Multimedia Appendix 2
Results of analyses excluding studies identified through backward citation searches.
[DOCX File, 24 KB - ojphi_v18i1e87559_app2.docx]

Multimedia Appendix 3
The full list of the 99 included studies and their primary characteristics.
[XLSX File, 110 KB - ojphi_v18i1e87559_app3.xlsx]

Multimedia Appendix 4
Time trend of consultation methods; Consultation methods by five-year period from 2006-2026 without backward citation.
[PNG File, 111 KB - ojphi_v18i1e87559_app4.png]

Checklist 1
PRISMA-ScR Checklist.
[DOCX File, 111 KB - ojphi_v18i1e87559_app5.docx]

References

1. WHO Guideline Recommendations on Digital Interventions for Health System Strengthening: World Health Organization; 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2026-06-29]
2. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 1;27(7):1132-1135. [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
3. Deldar K, Bahaadinbeigy K, Tara SM. Teleconsultation and clinical decision making: a systematic review. Acta Inform Med 2016 Jul 16;24(4):286-292. [doi: 10.5455/aim.2016.24.286-292] [Medline: 27708494]
4. Ye J, He L, Beestrum M. Implications for implementation and adoption of telehealth in developing countries: a systematic review of China's practices and experiences. NPJ Digit Med 2023 Sep 18;6(1):174. [doi: 10.1038/s41746-023-00908-6] [Medline: 37723237]
5. Scoping review for remote consultation from medical providers to specialists. Open Science Framework (OSF). URL: https://doi.org/10.17605/OSF.IO/K4TVU [accessed 2026-07-06]
6. 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. [doi: 10.1161/STROKEAHA.111.627943] [Medline: 21998052]
7. Otto CA, Shemenski R, Drudi L. Real-time tele-echocardiography: diagnosis and management of a pericardial effusion secondary to pericarditis at an Antarctic research station. Telemed J E Health 2012 Sep;18(7):521-524. [doi: 10.1089/tmj.2011.0266] [Medline: 22822790]
8. Jabarethina G, John KC, TanyaSruti I, Anitha R. Strategic competencies for telemedicine physicians: insights from a qualitative study on doctor-to-doctor healthcare delivery in India. BMC Health Serv Res 2025 Nov 18;25(1):1482. [doi: 10.1186/s12913-025-13564-1] [Medline: 41254728]
9. Álvarez-Dobaño JM, Toubes M, Novo-Platas J, et al. Cost-effectiveness of a new outpatient pulmonology care model based on physician-to-physician electronic consultation. Can Respir J 2022;2022:2423272. [doi: 10.1155/2022/2423272] [Medline: 36353446]

XSL•FO
RenderX

10. Olive MV, Gastaldi L, Corso M. Digitally-mediated coordination in healthcare: the effects of teleconsultation on doctor-to-doctor relational coordination. BMC Health Serv Res 2024 Feb 28;24(1):258. [doi: 10.1186/s12913-024-10726-5] [Medline: 38419009]
11. Markan A, Kishore A, Agarwal A, et al. Demographic profile of patients seeking teleophthalmology consultations through e-Sanjeevani: retrospective analysis of 5138 patients from North India. Indian J Ophthalmol 2022 Dec;70(12):4238-4243. [doi: 10.4103/ijo.IJO_781_22] [Medline: 36453323]
12. Pahalyants V, Murphy WS, Gunasekera NS, Das S, Hawryluk EB, Kroshinsky D. Evaluation of electronic consults for outpatient pediatric patients with dermatologic complaints. Pediatr Dermatol 2021 Sep;38(5):1210-1218. [doi: 10.1111/pde.14719] [Medline: 34467570]
13. Montelongo A, Becker JL, Roman R, et al. The management of COVID-19 cases through telemedicine in Brazil. PLOS ONE 2021;16(7):e0254339. [doi: 10.1371/journal.pone.0254339] [Medline: 34260644]
14. Tian PGJ, Harris JR, Seikaly H, Chambers T, Alvarado S, Eurich D. Characteristics and outcomes of physician-to-physician telephone consultation programs: environmental scan. JMIR Form Res 2021 Feb 23;5(2):e17672. [doi: 10.2196/17672] [Medline: 33620325]
15. Access to consultative dermatologic care via physician-to-physician asynchronous outpatient teledermatology. Am J Manag Care 2021 Jan 1;27(1):30-32. [doi: 10.37765/ajmc.2021.88574]
16. Karako K, Song P, Chen Y, Tang W. Realizing 5G- and AI-based doctor-to-doctor remote diagnosis: opportunities, challenges, and prospects. Biosci Trends 2020 Nov 4;14(5):314-317. [doi: 10.5582/bst.2020.03364] [Medline: 33100291]
17. Kong F, Horsham C, Rayner J, et al. Consumer preferences for skin cancer screening using mobile teledermoscopy: a qualitative study. Dermatology 2020;236(2):97-104. [doi: 10.1159/000505620] [Medline: 32126557]
18. Garber RN, Garcia E, Goodwin CW, Deeter LA. Pictures do influence the decision to transfer: outcomes of a telemedicine program serving an eight-state rural population. Journal of Burn Care & Research 2020 May 2;41(3):690-694. [doi: 10.1093/jbcr/iraa017]
19. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: randomized controlled trial. J Med Internet Res 2019 Feb 19;21(2):e11330. [doi: 10.2196/11330] [Medline: 30777845]
20. Azzolini C, Torreggiani A, Eandi C, et al. A teleconsultation network improves the efficacy of anti-VEGF therapy in retinal diseases. J Telemed Telecare 2013 Dec;19(8):437-442. [doi: 10.1177/1357633X13501760] [Medline: 24162839]
21. Hilty DM, Yellowlees PM, Cobb HC, Neufeld JD, Bourgeois JA. Use of secure e-mail and telephone: psychiatric consultations to accelerate rural health service delivery. Telemed J E Health 2006 Aug;12(4):490-495. [doi: 10.1089/tmj.2006.12.490] [Medline: 16942422]
22. Skladany L, Koller T, Adamcova Selcanova S, et al. Challenging management of severe chronic disorders in acute pandemic situation: chronic liver disease under COVID-19 pandemic as the proof-of-principle model to orchestrate the measures in 3PM context. EPMA J 2021 Mar;12(1):1-14. [doi: 10.1007/s13167-021-00231-8] [Medline: 33680218]
23. Clark AJ, Taenzer P, Drummond N, et al. Physician-to-physician telephone consultations for chronic pain patients: a pragmatic randomized trial. Pain Res Manag 2015;20(6):288-292. [doi: 10.1155/2015/345432] [Medline: 26474380]
24. Wong L, Martin-Khan M, Rowland J, Varghese P, Gray LC. The Rowland Universal Dementia Assessment Scale (RUDAS) as a reliable screening tool for dementia when administered via videoconferencing in elderly post-acute hospital patients. J Telemed Telecare 2012 Apr;18(3):176-179. [doi: 10.1258/jtt.2012.SFT113] [Medline: 22362836]
25. Ludwick DA, Lortie C, Doucette J, Rao J, Samoil-Schelstraete C. Evaluation of a telehealth clinic as a means to facilitate dermatologic consultation: pilot project to assess the efficiency and experience of teledermatology used in a primary care network. J Cutan Med Surg 2010;14(1):7-12. [doi: 10.2310/7750.2010.09012] [Medline: 20128984]
26. Ahmed Al-Kamel M. Feasibility of teledermatology in the eradication of leishmaniasis from Yemen project. Glob Dermatol ;4(2). [doi: 10.15761/GOD.1000206]
27. Kelsey A, Grant-Kels JM, Payette M. A high-tech ethical twist on the curbside consult. J Am Acad Dermatol 2015 Feb;72(2):349-351. [doi: 10.1016/j.jaad.2014.09.052] [Medline: 25592342]
28. Kaliyadan F, Ashique KT, Jagadeesan S, Krishna B. What's up dermatology? A pilot survey of the use of WhatsApp in dermatology practice and case discussion among members of WhatsApp dermatology groups. Indian J Dermatol Venereol Leprol 2016;82(1):67-69. [doi: 10.4103/0378-6323.171638] [Medline: 26728815]
29. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient reported outcomes with remote orthopaedic consultations by telemedicine: a randomised controlled trial. J Telemed Telecare 2019 Sep;25(8):451-459. [doi: 10.1177/1357633X18783921] [Medline: 29973130]
30. Prada C, Izquierdo N, Traipe R, Figueroa C. Results of a new telemedicine strategy in traumatology and orthopedics. Telemed J E Health 2020 May;26(5):665-670. [doi: 10.1089/tmj.2019.0090] [Medline: 31287786]
31. Talbot M, Harvey EJ, Berry GK, et al. A pilot study of surgical telementoring for leg fasciotomy. J R Army Med Corps 2018 May;164(2):83-86. [doi: 10.1136/jramc-2017-000817] [Medline: 29018173]
32. McKerrow M, North J. Clavicle fracture management: a comparison of a tertiary hospital and rural telehealth sites. J Telemed Telecare 2017 Dec;23(10):856-860. [doi: 10.1177/1357633X17733503] [Medline: 28958210]

XSL•FO
RenderX

33. Cota A, Tarchala M, Parent-Harvey C, et al. Review of 5.5 years' experience using e-mail-based telemedicine to deliver orthopedic care to remote communities. Telemed J E Health 2017 Jan;23(1):37-40. [doi: 10.1089/tmj.2016.0030] [Medline: 27336754]
34. Rowell PD, Pincus P, White M, Smith AC. Telehealth in paediatric orthopaedic surgery in Queensland: a 10-year review. ANZ J Surg 2014 Dec;84(12):955-959. [doi: 10.1111/ans.12753] [Medline: 25040240]
35. Ponce BA, Jennings JK, Clay TB, May MB, Huisingh C, Sheppard ED. Telementoring: use of augmented reality in orthopaedic education. Journal of Bone and Joint Surgery 2014;96(10):e84. [doi: 10.2106/JBJS.M.00928]
36. Daruwalla ZJ, Wong KL, Thambiah J. The application of telemedicine in orthopedic surgery in Singapore: a pilot study on a secure, mobile telehealth application and messaging platform. JMIR Mhealth Uhealth 2014 Jun 5;2(2):e28. [doi: 10.2196/mhealth.3303] [Medline: 25100283]
37. McGill A, North J. An analysis of an ongoing trial of rural videoconference fracture clinics. J Telemed Telecare 2012 Dec;18(8):470-472. [doi: 10.1258/jtt.2012.gth110] [Medline: 23209275]
38. Bertani A, Launay F, Candoni P, Mathieu L, Rongieras F, Chauvin F. Teleconsultation in paediatric orthopaedics in Djibouti: evaluation of response performance. Orthopaedics & Traumatology: Surgery & Research 2012 Nov;98(7):803-807. [doi: 10.1016/j.otsr.2012.03.022]
39. Morgan J, Walker S, Melaas D. Tele-orthopaedics: United States Army European Regional Medical Command. . 2012 p. 294-296. [doi: 10.3233/978-1-61499-022-2-294]
40. Blank E, Lappan C, Belmont PJ, et al. Early analysis of the United States Army's telemedicine orthopaedic consultation program. J Surg Orthop Adv 2011;20(1):50-55. [Medline: 21477534]
41. Hadjinicolaou MG, Nilavalan R, Itagaki T, Voskarides S, Pattichis CS, Schizas AN. Emergency TeleOrthoPaedics m-health system for wireless communication links. IET Commun 2009 Aug 10;3(8):1284-1296. [doi: 10.1049/iet-com.2008.0603]
42. Vladzymyrskyy AV. Our experience with telemedicine in traumatology and orthopedics. Ulus Travma Acil Cerrahi Derg 2004 Jul;10(3):189-191. [Medline: 15286891]
43. Ohinmaa A, Vuolio S, Haukipuro K, Winblad I. A cost-minimization analysis of orthopaedic consultations using videoconferencing in comparison with conventional consulting. J Telemed Telecare 2002;8(5):283-289. [doi: 10.1177/1357633X0200800507] [Medline: 12396857]
44. Harno K, Arajärvi E, Paavola T, Carlson C, Viikinkoski P. Clinical effectiveness and cost analysis of patient referral by videoconferencing in orthopaedics. J Telemed Telecare 2001;7(4):219-225. [doi: 10.1258/1357633011936435] [Medline: 11506757]
45. Haukipuro K, Ohinmaa A, Winblad I, Linden T, Vuolio S. The feasibility of telemedicine for orthopaedic outpatient clinics--a randomized controlled trial. J Telemed Telecare 2000;6(4):193-198. [doi: 10.1258/1357633001935347] [Medline: 11027118]
46. Lamminen H, Nevalainen J, Alho A, Lindholm TS, Tallroth K, Lepistö J. Experimental telemedicine in orthopaedics. J Telemed Telecare 1996;2(3):170-173. [doi: 10.1258/1357633961930013] [Medline: 9375053]
47. Barbieri JS, Nelson CA, James WD, et al. The reliability of teledermatology to triage inpatient dermatology consultations. JAMA Dermatol 2014 Apr;150(4):419-424. [doi: 10.1001/jamadermatol.2013.9517] [Medline: 24522374]
48. Lasierra N, Alesanco A, Gilaberte Y, Magallón R, García J. Lessons learned after a three-year store and forward teledermatology experience using internet: strengths and limitations. Int J Med Inform 2012 May;81(5):332-343. [doi: 10.1016/j.ijmedinf.2012.02.008] [Medline: 22425394]
49. Chen TS, Goldyne ME, Mathes EFD, Frieden IJ, Gilliam AE. Pediatric teledermatology: observations based on 429 consults. J Am Acad Dermatol 2010 Jan;62(1):61-66. [doi: 10.1016/j.jaad.2009.05.039] [Medline: 19926163]
50. Philp JC, Frieden IJ, Cordoro KM. Pediatric teledermatology consultations: relationship between provided data and diagnosis. Pediatr Dermatol 2013;30(5):561-567. [doi: 10.1111/pde.12180] [Medline: 23889079]
51. Gehris RP. Pediatric teledermatology: improve access in an academic children's hospital. American Telemedicine Association 2018 [FREE Full text]
52. Warshaw EM, Gravely AA, Nelson DB. Reliability of store and forward teledermatology for skin neoplasms. J Am Acad Dermatol 2015 Mar;72(3):426-435. [doi: 10.1016/j.jaad.2014.11.001] [Medline: 25599624]
53. Holmes AN, Chansky PB, Simpson CL. Teledermatology consultation can optimize treatment of cutaneous disease by nondermatologists in under-resourced clinics. Telemed J E Health 2020 Oct;26(10):1284-1290. [doi: 10.1089/tmj.2019.0217] [Medline: 31800369]
54. Armstrong IJ, Haston WS. Medical decision support for remote general practitioners using telemedicine. J Telemed Telecare 1997;3(1):27-34. [doi: 10.1258/1357633971930166] [Medline: 9139758]
55. Rogers FB, Ricci M, Caputo M, et al. The use of telemedicine for real-time video consultation between trauma center and community hospital in a rural setting improves early trauma care: preliminary results. Journal of Trauma and Acute Care Surgery 2001 Dec;51(6):1037-1041. [doi: 10.1097/00005373-200112000-00002]
56. Agarwal AK, Gaieski DF, Perman SM, et al. Telemedicine REsuscitation and Arrest Trial (TREAT): a feasibility study of real-time provider-to-provider telemedicine for the care of critically ill patients. Heliyon 2016 Apr;2(4):e00099. [doi: 10.1016/j.heliyon.2016.e00099] [Medline: 27441272]

XSL•FO
RenderX

57. Gerhardt RT, Berry J, Mabry RL, et al. Evaluation of contingency telemedical support to improve casualty care at a simulated military intermediate resuscitation facility: the EM-ANGEL study. J Spec Oper Med 2014;14(1):50-57. [doi: 10.55460/G0D7-OBXG] [Medline: 24604439]
58. Agarwal S, Day DJ, Sibson L, et al. Thrombolysis delivery by a regional telestroke network--experience from the U.K. National Health Service. J Am Heart Assoc 2014 Feb 26;3(1):e000408. [doi: 10.1161/JAHA.113.000408] [Medline: 24572251]
59. Engle X, Aird J, Tho L, et al. Combining continuing education with expert consultation via telemedicine in Cambodia. Trop Doct 2014 Apr;44(2):62-68. [doi: 10.1177/0049475513515654] [Medline: 24322763]
60. Seeber A, Mitterer M, Gunsilius E, et al. Feasibility of a multdisciplinary lung cancer videoconference between a peripheral hospital and a comprehensive cancer centre. Oncology 2013;84(3):186-190. [doi: 10.1159/000345314] [Medline: 23328311]
61. Akkoyun I. The advantages of using photographs and video images in telephone consultations with a specialist in paediatric surgery. Afr J Paediatr Surg 2012;9(2):128-131. [doi: 10.4103/0189-6725.99398] [Medline: 22878761]
62. Walsh C, Cosgrave J, Crean P, et al. Synchronized, interactive teleconferencing with digital cardiac images. J Digit Imaging 2006 Mar;19(1):85-91. [doi: 10.1007/s10278-005-8147-0] [Medline: 16249837]
63. Greiner A, Mess WH, Schmidli J, et al. Cyber medicine enables remote neuromonitoring during aortic surgery. J Vasc Surg 2012 May;55(5):1227-1232. [doi: 10.1016/j.jvs.2011.11.121] [Medline: 22341581]
64. Tamariz F, Merrell R, Popescu I, et al. Design and implementation of a web-based system for intraoperative consultation. World J Surg 2009 Mar;33(3):448-454. [doi: 10.1007/s00268-008-9858-4] [Medline: 19123027]
65. Ryskal O, Muschinskaja M, Göbel U, et al. Telemicroscopic conferences for children of the perm territory with suspected or proven malignant solid tumors. Klin Padiatr 2010 May;222(3):199-202. [doi: 10.1055/s-0030-1251987] [Medline: 20514632]
66. Froehlich W, Seitaboth S, Chanpheaktra N, Pugatch D. Case report: an example of international telemedicine success. J Telemed Telecare 2009;15(4):208-210. [doi: 10.1258/jtt.2008.081001] [Medline: 19471034]
67. Ohno G, Watanabe K, Okada Y, Higuchi K. Practical experience of telehealth between an Antarctic station and Japan. J Telemed Telecare 2012 Dec;18(8):473-475. [doi: 10.1258/jtt.2012.gth111] [Medline: 23209273]
68. Otto C, Shemenski R, Scott JM, Hartshorn J, Bishop S, Viegas S. Evaluation of tele-ultrasound as a tool in remote diagnosis and clinical management at the Amundsen-Scott South Pole Station and the McMurdo Research Station. Telemed J E Health 2013 Mar;19(3):186-191. [doi: 10.1089/tmj.2012.0111] [Medline: 23480714]
69. Callahan CW, Malone F, Estroff D, Person DA. Effectiveness of an Internet-based store-and-forward telemedicine system for pediatric subspecialty consultation. Arch Pediatr Adolesc Med 2005 Apr;159(4):389-393. [doi: 10.1001/archpedi.159.4.389] [Medline: 15809396]
70. Stutchfield BM, Jagilly R, Tulloh BR. Second opinions in remote surgical practice using email and digital photography. ANZ J Surg 2007 Nov;77(11):1009-1012. [doi: 10.1111/j.1445-2197.2007.04299.x] [Medline: 17931268]
71. Syed-Abdul S, Scholl J, Chen CC, et al. Telemedicine utilization to support the management of the burns treatment involving patient pathways in both developed and developing countries. Journal of Burn Care & Research 2012;33(4):e207-e212. [doi: 10.1097/BCR.0b013e318241b6b7]
72. Costanzo G, Monari P. Experience with asynchronous medical teleconsultation in the alliance of the Italian hospitals worldwide. J Telemed Telecare 2006;12(8):382-386. [doi: 10.1258/135763306779378645] [Medline: 17227601]
73. Zolfo M, Bateganya MH, Adetifa IM, Colebunders R, Lynen L. A telemedicine service for HIV/AIDS physicians working in developing countries. J Telemed Telecare 2011;17(2):65-70. [doi: 10.1258/jtt.2010.100308] [Medline: 21078680]
74. Maddry JK, Sessions D, Heard K, Lappan C, McManus J, Bebarta VS. Wartime toxicology: evaluation of a military medical toxicology telemedicine consults service to assist physicians serving overseas and in combat (2005-2012). J Med Toxicol 2014 Sep;10(3):261-265. [doi: 10.1007/s13181-014-0398-z] [Medline: 24752493]
75. McManus J, Salinas J, Morton M, Lappan C, Poropatich R. Teleconsultation program for deployed soldiers and healthcare professionals in remote and austere environments. Prehosp Disaster Med 2008;23(3):210-216. [Medline: 18702266]
76. Mines MJ, Bower KS, Lappan CM, Mazzoli RA, Poropatich RK. The United States army ocular teleconsultation program 2004 through 2009. Am J Ophthalmol 2011 Jul;152(1):126-132. [doi: 10.1016/j.ajo.2011.01.028] [Medline: 21570049]
77. Rand ER, Lappan CM, Lee JC, Furgerson JL. Paging the worldwide cardiology consultant: the army knowledge online telemedicine consultation program in cardiology. Mil Med 2009 Nov;174(11):1144-1148. [doi: 10.7205/milmed-d-00-2309] [Medline: 19960820]
78. Schmidt T, Lappan CM, Hospenthal DR, Murray CK. Deployed provider satisfaction with infectious disease teleconsulation. Mil Med 2011 Dec;176(12):1417-1420. [doi: 10.7205/milmed-d-11-00237] [Medline: 22338358]
79. Waterman BR, Laughlin MD, Belmont PJ, Schoenfeld AJ, Pallis MP. Enhanced casualty care from a global military orthopaedic teleconsultation program. Injury 2014 Nov;45(11):1736-1740. [doi: 10.1016/j.injury.2014.03.012] [Medline: 24810665]
80. Kadaba V, Ly T, Noor S, et al. A hybrid approach to telepathology in Cambodia. J Telemed Telecare 2013 Dec;19(8):475-478. [doi: 10.1177/1357633X13512071] [Medline: 24197402]
81. Helveston EM, Neely DE, Cherwek DH, Smallwood LM. Diagnosis and management of strabismus using telemedicine. Telemed J E Health 2008 Aug;14(6):531-538. [doi: 10.1089/tmj.2007.0086] [Medline: 18729751]

XSL•FO
RenderX

82. Hu SW, Foong HBB, Elpern DJ. Virtual grand rounds in dermatology: an 8 - year experience in web - based teledermatology. Int J Dermatol 2009 Dec;48(12):1313-1319. [doi: 10.1111/j.1365-4632.2009.04112.x] [Medline: 20415672]
83. Massone C, Soyer HP, Hofmann-Wellenhof R, et al. Two years' experience with web-based teleconsulting in dermatology. J Telemed Telecare 2006;12(2):83-87. [doi: 10.1258/135763306776084419] [Medline: 16539755]
84. Patterson V, Wootton R. A web-based telemedicine system for low-resource settings 13 years on: insights from referrers and specialists. Glob Health Action 2013 Sep 23;6(1):21465. [doi: 10.3402/gha.v6i0.21465] [Medline: 24063752]
85. Börve A, Dahlén Gyllencreutz J, Terstappen K, et al. Smartphone teledermoscopy referrals: a novel process for improved triage of skin cancer patients. Acta Derm Venereol 2015 Feb;95(2):186-190. [doi: 10.2340/00015555-1906] [Medline: 24923283]
86. Chigusa Y, Imatake H, Komatsu M, Ikeda Y, Mogami H, Mandai M. Implementing a doctor-to-doctor telemedicine network to strengthen rural obstetric care in Japan: feasibility, clinical utility, and educational impact. Cureus 2026 Jan;18(1):e102110. [doi: 10.7759/cureus.102110] [Medline: 41732641]
87. Heimbuch S, Tischler L, Beyer A, et al. Telemedicine in paediatrics: acceptance and satisfaction from parents' perspective [Article in German]. Gesundheitswesen 2025 Dec;87(S 03):S317-S323. [doi: 10.1055/a-2543-3179] [Medline: 40389223]
88. Ghasemi S, Changiz T, Omid A. Essential competencies for an effective physician-to-physician teleconsultation. BMC Med Educ 2025 Nov 25;26(1):11. [doi: 10.1186/s12909-025-08362-6] [Medline: 41291678]
89. Yan YAN, Naito T. Policies and practices of cross-border teleconsultation in Japan. Juntendo Med J 2025;71(6):385-388. [doi: 10.14789/ejmj.JMJ25-0016-P] [Medline: 41608220]
90. Kudo K, Ueda S, Hisada Y, et al. Training needs analysis on the conducting doctor to doctor telemedicine conference/consultations in Asian medical institutions. Stud Health Technol Inform 2025 Aug 7;329:1437-1441. [doi: 10.3233/SHTI251076] [Medline: 40776094]
91. Kudo K, Hisada Y, Ueda S, Kikukawa M, Nakashima N, Moriyama T. Inter-country doctor-to-doctor telemedicine conferences/consultations following COVID-19: a Survey of the National University Hospital Council of Japan. Telemed Rep 2025;6(1):148-161. [doi: 10.1089/tmr.2025.0015] [Medline: 40630803]
92. Fagan M, Levano S, Haughton J, et al. A rural ambulance service in Northern Togo improves access to emergency care for women with obstetric complications. J Glob Health ;14. [doi: 10.7189/jogh.14.04201]
93. Sherif NA, Lorch AC, Armstrong GW. Physician-to-physician eConsultations to ophthalmologists at an academic medical center. Transl Vis Sci Technol 2024 Oct 1;13(10):13. [doi: 10.1167/tvst.13.10.13] [Medline: 39377754]
94. Niknam F, Mardani M, Bastani P, et al. Assessing the usability and reliability of a web-based teledentistry tool for remote diagnosis of oral lesions: a cross-sectional study. BMC Oral Health 2024 Sep 16;24(1):1094. [doi: 10.1186/s12903-024-04696-7] [Medline: 39285380]
95. Joshi K, Modi B, Katoch CDS. Utilization of telemedicine services of institute of national importance in the western region of India: a mixed-method study. J Family Med Prim Care 2024 Sep;13(9):3782-3787. [doi: 10.4103/jfmpc.jfmpc_56_24] [Medline: 39464921]
96. Shimizu E, Kamezaki M, Nishimura H, Nakayama S, Toda I. A Case of traumatic hyphema diagnoses by telemedicine between a remote island and the mainland of Tokyo. Cureus 2024 Jul;16(7):e65153. [doi: 10.7759/cureus.65153] [Medline: 39176324]
97. Goldberg-Stein S, Varghese J, Wang JJ, et al. Global health initiatives: international physician-to-physician consultation programs. J Am Coll Radiol 2024 Aug;21(8):1239-1247. [doi: 10.1016/j.jacr.2024.04.023] [Medline: 38763443]
98. Klingberg A, Wallis LA, Hasselberg M, Yen PY, Fritzell SC. Teleconsultation using mobile phones for diagnosis and acute care of burn injuries among emergency physicians: mixed-methods study. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e11076. [doi: 10.2196/11076] [Medline: 30341047]
99. Kasie MN, Guadie HA, Gashu KD. Mobile phone-based teleconsultation service utilization and associated factors among healthcare professionals in primary hospitals of West Gojjam zone, Northwest Ethiopia: a cross-sectional study. In Review . [doi: 10.21203/rs.3.rs-601821/v1]
100. Choemprayong S, Charoenlap C, Piromsopa K. Exploring usability issues of a smartphone-based physician-to-physician teleconsultation app in an orthopedic clinic: mixed methods study. JMIR Hum Factors 2021 Dec 20;8(4):e31130. [doi: 10.2196/31130] [Medline: 34931991]
101. Gonzalez BE, Sabella C, Esper FP, et al. Physician-to-physician electronic consultation: a tool for the pediatric infectious diseases specialist to document encounters and quantify effort. J Pediatric Infect Dis Soc 2021 Apr 3;10(3):334-336. [doi: 10.1093/jpids/piaa041] [Medline: 32369170]
102. Imaizumi H, Watanabe A, Hirano H, Takemura M, Kashiwagi H, Monobe S. Hippocra: doctor-to-doctor teledermatology consultation service towards future AI-based diagnosis system in Japan. 2017 Presented at: 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW); Jun 12-14, 2017. [doi: 10.1109/ICCE-China.2017.7990990]
103. Li Y, Wong A. Patients' and family caregivers' perceptions of doctor-to-doctor patients' and family caregivers' perceptions of doctor-to-doctor advice and electronic referral notifications in Alberta. Patient Exp J 2021;8(1):77-87 [FREE Full text] [doi: 10.35680/2372-0247.1441]

XSL•FO
RenderX

104. Muse ED, Godino JG, Netting JF, Alexander JF, Moran HJ, Topol EJ. From second to hundredth opinion in medicine: a global consultation platform for physicians. NPJ Digit Med 2018;1(1):55. [doi: 10.1038/s41746-018-0064-y] [Medline: 31304334]
105. Pérez-Sánchez S, Quesada García M, Sánchez Fernández F, et al. Neurological primary care: assessment of teleconsultation as the only access gateway to neurological care from primary care. Neurologia (Engl Ed) 2026 Apr;41(3):501919. [doi: 10.1016/j.nrleng.2026.501919] [Medline: 41765133]
106. Mathews D, Abernethy A, Butte AJ, et al. Telehealth and mobile health: case study for understanding and anticipating emerging science and technology. NAM Perspect 2023;2023. [doi: 10.31478/202311e] [Medline: 38812841]
107. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
108. Lo S, Rose A, Fowers S, et al. Ghana 3D telemedicine international MDT: a proof-of-concept study. J Plast Reconstr Aesthet Surg 2024 Jan;88:425-435. [doi: 10.1016/j.bjps.2023.10.130] [Medline: 38091684]
109. Lo S, Fowers S, Darko K, et al. Participatory development of a 3D telemedicine system during COVID: the future of remote consultations. J Plast Reconstr Aesthet Surg 2023 Dec;87:479-490. [doi: 10.1016/j.bjps.2022.10.012] [Medline: 36890078]
110. Wu CT, Lin TY, Lin CJ, Hwang DK. The future application of artificial intelligence and telemedicine in the retina: a perspective. Taiwan J Ophthalmol 2023;13(2):133-141. [doi: 10.4103/tjo.TJO-D-23-00028] [Medline: 37484624]
111. Garcia JP, Avila FR, Torres-Guzman RA, et al. A narrative review of telemedicine and its adoption across specialties. Mhealth 2024;10:19. [doi: 10.21037/mhealth-23-28] [Medline: 38689613]
112. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation in telemedicine use and outpatient care during the COVID-19 pandemic in the United States. Health Aff (Millwood) 2021 Feb;40(2):349-358. [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
113. Gaur P, Ganguly AP, Kuo M, et al. Childcare needs as a barrier to healthcare among women in a safety-net health system. BMC Public Health 2024 Jun 17;24(1):1608. [doi: 10.1186/s12889-024-19125-1] [Medline: 38886715]
114. Hatef E, Wilson RF, Zhang A, et al. Effectiveness of telehealth versus in-person care during the COVID-19 pandemic: a systematic review. NPJ Digit Med 2024 Jun 15;7(1):157. [doi: 10.1038/s41746-024-01152-2] [Medline: 38879682]
115. Sugawara Y, Hirakawa Y, Nangaku M. Telemedicine in nephrology: future perspective and solutions. Clin Kidney J 2024 Dec;17(Suppl 2):1-8. [doi: 10.1093/ckj/sfae267] [Medline: 39583140]
116. Li Q, Cheng F, Zeng H, Xu J. Health insurance payment for telehealth services: scoping review and narrative synthesis. J Med Internet Res 2024 Dec 9;26:e56699. [doi: 10.2196/56699] [Medline: 39652868]
117. Global Strategy on Digital Health 2020–2025: World Health Organization; 2021. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2026-06-29]

Abbreviations

CPT: current procedural terminology
HIC: high-income country
LMIC: low- and middle-income country
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by E Mensah; submitted 02.Dec.2025; peer-reviewed by S Kinoshita, V Palama; revised version received 07.May.2026; accepted 06.Jun.2026; published 15.Jul.2026.

Please cite as:
Hara R, Hirakawa Y, Sugawara Y, Iwagami M, Inokuchi R, Nangaku M
Trend in Devices and Digital Tools for Remote Consultation From Medical Providers to Specialists: Scoping Review
Online J Public Health Inform 2026;18:e87559
URL: https://ojphi.jmir.org/2026/1/e87559
doi:10.2196/87559

© Risa Hara, Yosuke Hirakawa, Yuka Sugawara, Masao Iwagami, Ryota Inokuchi, Masaomi Nangaku. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 15.Jul.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adaptation of Social Health Care Tools From Paper-Based to Web-Based Formats: Scoping Review

Charlèss Dupont[1], PhD; Eva Savolainen[2], PhD; Petter Fjällström[3], PhD; Malin Eneslätt[2,4], PhD

[1]End-of-Life Care Research Group, Vrije Universiteit Brussel, Brussels, Belgium

[2]Department of Health, Education and Technology, Luleå University of Technology, Universitetsområdet, Luleå, Norrbotten, Sweden

[3]Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Västerbotten, Sweden

[4]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Stockholm, Sweden

Corresponding Author:
Malin Eneslätt, PhD
Department of Health, Education and Technology, Luleå University of Technology, Universitetsområdet, Luleå, Norrbotten, Sweden

Abstract

Background: The transition of social health care tools from paper-based to web-based formats is increasingly common in social and health care settings. While digital delivery may improve accessibility, flexibility, and scalability, such transitions require careful adaptation of content, design, and workflows. However, evidence on how these transitions are carried out in practice, and which challenges and strategies are involved, remains fragmented.

Objective: This scoping review aimed to map how social health care tools have transitioned from paper-based to web-based formats and to synthesize evidence on tool characteristics and target groups, adaptations made during the transition, strategies and methodologies used, and reported challenges and solutions.

Methods: A scoping review was conducted following the Arksey and O'Malley framework. Systematic searches were performed in PubMed, PsycINFO, Scopus, and Web of Science to identify English-language empirical studies describing the transition of social health care tools from paper-based to web-based formats. Social health care tools were defined as technologies supporting communication, reflection, or informed decision-making among patients, family caregivers, and/or health care professionals. Two reviewers independently screened studies and extracted data using a standardized form. Targeted gray literature searches were conducted to supplement missing information about the included tools. Findings were synthesized narratively and presented in tables.

Results: Eleven tools were included. Most were delivered as web-based platforms (n=8), with additional tools delivered as mobile apps (n=2) or in a game-based format (n=1). The tools targeted diverse populations, including people with chronic or life-limiting conditions, family caregivers, adolescents, parents, pregnant women, and health care professionals. Reported reasons for transitioning to digital formats included improving accessibility and reach (n=7), enhancing user engagement through interactive features (n=5), and increasing efficiency or scalability (n=3). Across tools, core content was largely retained, while adaptations mainly focused on format and delivery, such as multimedia elements, interactive modules, tailored feedback, and usability improvements. User-centered and participatory design approaches were commonly reported, sometimes combined with iterative or agile development methods. Common challenges included usability issues, low engagement, technical integration difficulties, administrative delays, and resource constraints, with solutions focusing on iterative redesign, clearer instructions, increased interactivity, and hybrid paper-digital approaches.

Conclusions: This scoping review examines the process of transitioning social health care tools from paper-based to web-based formats, rather than focusing solely on effectiveness outcomes. It synthesizes how adaptations are made, which strategies are used, and which challenges arise during digitalization. The findings identify common adaptation patterns and gaps in reporting, particularly regarding target group–specific design and equity considerations. These insights may support developers, researchers, and policymakers in planning more usable, inclusive, and context-sensitive digital transitions in social and health care settings.

(Online J Public Health Inform 2026;18:e88631) doi:10.2196/88631

KEYWORDS

digital health; user-centered design; implementation; scoping review; eHealth; usability

XSL•FO
RenderX

Introduction

In recent years, a notable paradigm shift has occurred with the transition of tools in social and health care settings from traditional paper-based to interactive online formats [1,2]. For instance, digital tools now replace paper diaries, allowing patients to enter their medical information from home in real time [3]. This enhances data accuracy and timeliness, enabling health care providers to offer more personalized and proactive care [1,4].

This transition takes place within a broader process of digital transformation in health and social care systems. Recent literature emphasizes that digitalization is not only about introducing new technologies but also about adapting existing tools and care processes to digital environments [5,6]. However, while much research has focused on the development and evaluation of new digital interventions, less attention has been paid to how existing paper-based tools are systematically transitioned to web-based formats. It is crucial to recognize that this shift necessitates careful adaptation to ensure that online tools are perceived and used effectively [7,8]. Unlike paper-based tools, online tools may present information differently and require users to navigate through various digital interfaces [7].

Despite the advantages, such as improved accessibility and real-time updates, transitioning from paper to digital formats may present several challenges. These include user resistance, digital literacy issues, accessibility barriers, and the need for technical support and training [9,10]. Recent studies have further highlighted ongoing digital inequalities in health care access, particularly among older adults and socially vulnerable groups [11-13].

Individual studies have reported mixed results regarding the effectiveness of transitioning from paper-based to online formats [14,15]. In some cases, digital tools are designed to closely resemble their paper versions to support familiarity. However, this approach may not always be suitable in an online environment or for different target groups. Importantly, the transition from paper to web-based formats is not merely a technical conversion; it often involves conceptual and structural changes that may affect usability, user engagement, and accessibility [16]. Moreover, these studies typically evaluate outcomes after implementation and provide limited insight into how and why specific adaptations were made during the transition process.

There is still limited knowledge about how existing tools are adapted for online use, which specific modifications are made during this process, and how these adaptations influence their usability and accessibility. To our knowledge, no comprehensive overview has systematically mapped which tools in social health care settings have undergone such transitions, how these transitions were performed, and what challenges and solutions were reported. A clearer understanding of these processes is needed to better inform future transitions and to avoid repeating common design or implementation problems.

Therefore, this paper explores how tools used in social and health care settings (hereafter called social health care tools) have been transitioned from paper-based to web-based formats. It examines their characteristics, strategies, methodologies, challenges, and solutions:

1. Which social health care tools have shifted from paper-based to web-based formats, and what are their target groups, countries of development, formats, aims, and the reasons behind the transition?
2. What specific changes were made when transitioning social health care tools from traditional paper-based formats to web-based platforms?
3. What strategies and methodologies were used to transition the tools from paper-based formats to online platforms?
4. What were the key challenges encountered during the transition, and how were these challenges addressed?

Methods

Design and Frameworks

Scoping review studies map key concepts and evidence in a research area and can stand alone, particularly in complex or underreviewed domains [17]. For our scoping review focusing on studies or reports on transitioning social health care tools from traditional paper-based to web-based formats, we adopted the methodological framework outlined by Arksey and O'Malley [17], with no critical appraisal of the studies. This framework involves 5 stages: identifying the research question, locating relevant studies, selecting studies, charting the data, and collating and summarizing the results. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist guided the reporting of this scoping review (Checklist 1) [18]. In addition, the search strategy and reporting were documented according to the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension) extension to strengthen transparency and reproducibility of each component of the search [18]. No prior protocol was established or published.

Eligibility Criteria

Eligibility criteria were defined using the Population-Concept-Context framework recommended for scoping reviews [19]. Studies were eligible if they reported on the following criteria:

1. Social health care tools used by patients, family caregivers, health care professionals, or related stakeholders (Population).
2. The transition of a tool from a paper-based to a web-based or internet-based format (Concept).

Social health care tools were defined as structured instruments intended to support communication, reflection, informed decision-making, self-management, coordination, or psychosocial support within social or health care settings.

Empirical qualitative, quantitative, or mixed methods studies conducted in any clinical, community, or organizational context were eligible (Context), provided they reported on adaptations

XSL•FO
RenderX

made during the transition process and/or strategies, methodologies, challenges, or solutions related to digitalization.

Studies were excluded if they did not involve information on the paper-to-web transition, focused exclusively on physical health without a social or communicative component, were not original empirical research, or described transitions in tools delivered via non–internet-based formats (eg, CD-ROM only). Only English-language publications were included. Gray literature was consulted solely to supplement missing information on included tools.

Information Sources and Search Strategies

To ensure a thorough exploration of the relevant literature, we used a comprehensive search strategy. This strategy involved systematic searches of electronic databases (Multimedia Appendix 1), including PubMed, PsycINFO, Scopus, and Web of Science, to identify peer-reviewed articles and scholarly publications. The full electronic search strategies for each database, including all search terms, Boolean operators, applied limits, and the date on which each search was performed, are provided in Multimedia Appendix 1 in accordance with PRISMA-S [18]. Searches were limited to English-language publications.

In addition, we supplemented the database searches with backward citation searching, consistent with Joanna Briggs Institute recommendations for scoping reviews [20]. After full-text screening, the reference lists of all included articles were examined to identify additional potentially relevant sources describing the transition of social health care tools from paper-based to web-based formats. Records identified through this process were retrieved and assessed against the same eligibility criteria as database-derived records. Eligible studies identified through citation searching were included in the review and integrated into data extraction and synthesis.

Additionally, recognizing the potential significance of nonacademic sources, we searched for extra information on the included tools in gray literature, such as conference proceedings, reports, theses, and dissertations. Two researchers (CD and ES) independently conducted the searches, using the name of the included tool or other identifiers (author names, institution names, etc). Each researcher conducted the searches on separate computers, ensuring privacy by clearing browser history and cookies and avoiding logins to Google accounts. The content of the search results, such as executive summaries, was examined to determine missing information on adapting a web-based support tool, such as the adaptation process and challenges encountered during the transition.

Selection of Sources

First, CD searched all aforementioned databases and screened records based on their titles. All records retrieved from the database searches were imported into an Excel file, where duplicates were removed, and the deduplicated records were imported to Rayyan software. After deduplication, 2 reviewers (CD and ES) independently screened the titles and abstracts of retrieved articles for relevance to the research questions and eligibility criteria. Full texts of potentially relevant articles were assessed for final inclusion. Any discrepancies in screening and selection were resolved through discussion and consensus with all authors. Records identified through backward citation searching were screened using the same process.

For each included tool, targeted gray literature searches (eg, via Google) were performed using the tool name, author names, or institutional affiliations to retrieve additional publicly available information on the adaptation process. These searches were used solely to supplement missing details when needed and did not lead to the inclusion of additional tools.

Data Charting, Items, and Synthesis

Data extraction was conducted independently by 2 researchers (CD and ES) using a standardized data extraction form. The form was piloted and refined before full extraction. Data were extracted to address the 4 research questions, including study and tool characteristics, target groups, reasons for transition, adaptations made during the transition, development methodologies and strategies, and reported challenges and solutions. Where relevant, targeted gray literature searches were used to supplement missing information on included tools. Extracted data were synthesized per research question and presented narratively and in tables.

Although eligibility screening was performed at the level of published papers, synthesis was conducted at the level of the tool. This approach was chosen because multiple publications may describe the same tool, and the aim of this review was to analyze tool characteristics and adaptation processes rather than publication outcomes.

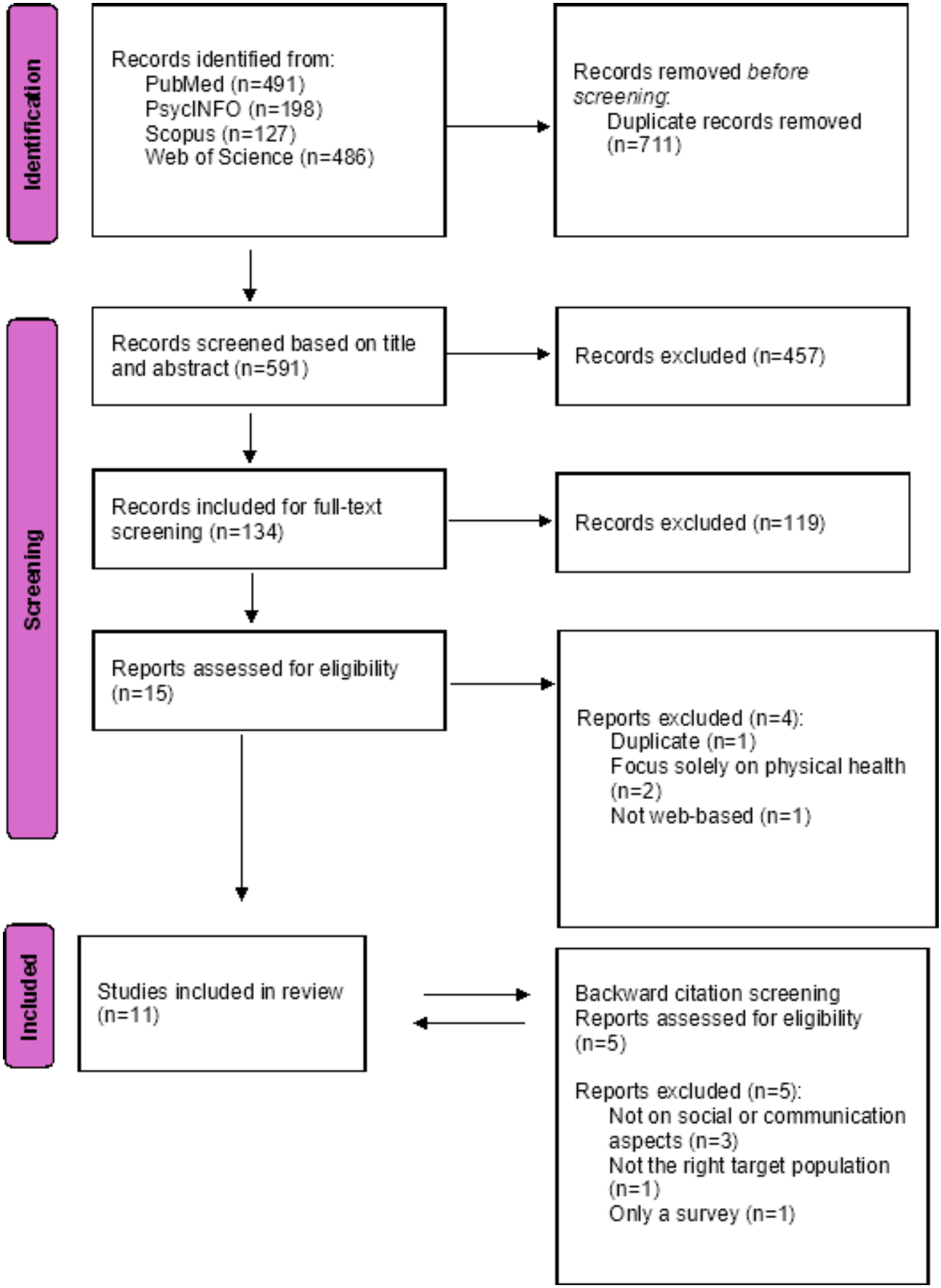
Results

Selection Process

The search across academic databases resulted in a total of 1302 peer-reviewed articles: 491 from PubMed, 486 from Web of Science, 198 from PsycINFO, and 127 from Scopus. All results were imported into an Excel file to remove duplicates. After removing duplicates, CD screened the records based on titles, and 591 articles remained. These articles were screened by CD and ES based on abstracts, after which 134 articles were selected for full-text review. Based on this full-text screening, 15 tools were initially included. However, following a discussion among all authors on the eligibility criteria and the included tools, 4 of these tools were excluded. One was a duplicate already included from another source, 2 focused exclusively on physical health and did not address social or communicative aspects of health care, and 1 tool was delivered via CD-ROM rather than a web-based format. No additional tools were included based on backward citation searching. An overview of the selection process is shown in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection process.



XSL•FO
RenderX

Characteristics of Social Health Care Tools Transitioned From Paper-Based to Web-Based Formats

Table 1 provides an overview of the 11 included social health care tools (hereafter called tools), including their aims, formats, target groups, countries of development, and reported reasons for the transition from paper-based to web-based delivery [21-31]. Most tools were developed as web-based platforms (n=8; [22,23,25-27,30,31]), with newer tools delivered as mobile apps (n=2; [24,29]) or in a game-based format (n=1; [21]).

Table . Characteristics of social health care tools transitioned from paper-based to web-based formats.

Study	Tool name	Target population	Aim of the tool	Country of development	Digital format	Reason for transition
Bayley and Brown [21], 2015	What Should We Tell the Children? (WSWTTC)	Parents of children and adolescents	Support parent-child communication about sex and relationships	United Kingdom	Game-based	Reduce participation barriers and increase engagement
Honary et al [22], 2018	Relatives Education And Coping Toolkit (REACT)	Relatives of people with psychosis or bipolar disorder	Provide informational and emotional support	United Kingdom	Web-based	Improve accessibility and scalability
Kingston et al [23], 2015	Not specified	Pregnant women	Support mental health screening	Canada	Web-based	Improve efficiency and reduce disclosure barriers
Lee et al [24], 2023	Promoting Independence in Dementia (PRIDE) app	People with mild dementia	Support self-management and independence	United Kingdom	App	Improve usability and accessibility
Monnet et al [25], 2023	Levenswensen kaarten	People with dementia and family caregivers	Support advance care planning	Belgium	Web-based	Improve accessibility and usability
Peels et al [26], 2012	Active Plus	Adults aged ≥50 years	Promote physical activity	Netherlands	Web-based	Improve reach and enable tailoring at scale
Staffileno et al [27], 2015	Not specified	Young African American women	Promote lifestyle behavior change	United States	Web-based	Improve accessibility and engagement
Stille et al [28], 2015	Parent-Provider Partnerships in Referral Communication (P3RC)	Health care professionals and parents	Improve referral communication	United States (Colorado, Oklahoma)	Web-based	Improve workflow integration and efficiency
Svarre et al [29], 2017	Housing Enabler app	Health care professionals	Assess environmental accessibility	Denmark	App	Improve data accuracy and efficiency
Van Goethem et al [30], 2023	Web-based program for Family involvement, Optimistic outlook, Coping effectiveness, Uncertainty reduction, and Symptom management (iFOCUS)	People with advanced cancer and family caregivers	Support emotional well-being and self-efficacy	Belgium, Netherlands, Denmark, Italy, United Kingdom, and Ireland	Web-based	Improve accessibility and self-management
Whitehouse et al [31], 2013	TickiT	Adolescents in health care settings	Support psychosocial screening	Canada	Web-based/App	Improve efficiency and youth engagement

The tools addressed a range of aims within social and health care settings, with most focusing on psychosocial support, self-management, communication, or screening. Target groups included patients with chronic or life-limiting conditions, family caregivers, adolescents, parents, pregnant women, and health care professionals. While several tools were primarily intended for patients or caregivers, a smaller subset targeted health care professionals and focused on clinical or administrative processes.

Tools were developed across multiple countries, predominantly in Europe and North America. Across studies, the most frequently reported reasons for transitioning to digital formats were to improve accessibility and availability (eg, enabling use from home and outside clinical appointments), enhance user engagement through interactive features (eg, videos, quizzes, tailored feedback, automated scoring, or reminders), and increase efficiency or scalability. In some cases, external contextual factors, such as the COVID-19 pandemic, were

XSL•FO
RenderX

reported as contributing to the decision to digitalize existing tools.

In general, information on the transition process was limited. Information provided described the transition from paper-based to web-based formats as an approach to address limitations of paper tools, such as limited opportunities for remote use, reduced flexibility, or challenges in routine implementation, while preserving the core purpose and content of the original tools.

Changes Made During the Transition

A range of adaptations were made during the transition from paper-based to web-based formats, tailored to the function, delivery setting, and target audience of each intervention (see Table 2). In 5 tools [21,22,24,26,27], the original intervention content was largely retained, while the format and delivery were modified to include digital features such as videos, online modules, quizzes, or interactive exercises. In 3 tools [21,22,30], the structure and delivery of sessions were also redesigned, for example, moving from face-to-face or nurse-led delivery to self-guided or asynchronous modules [30]. Five tools [21,22,25,27,31] integrated features that allowed users to navigate content adaptively or receive tailored feedback, such as adaptive navigation, customized feedback, or emotionally responsive content. Design changes to improve usability, such as font adjustments, simplified navigation, and tailored login systems, were also reported.

XSL•FO
RenderX

Table . Overview of adaptations made during the transition from paper-based to web-based formats.

Study	Original paper-based feature	Web-based adaptation and purpose of change
Bayley and Brown [21], 2015	"What Should We Tell the Children?" group program with exercises to support parent-child communication about sex and relationships.	Converted into an online serious game with first-person role-play, interactive quizzes, and feedback. Purpose: to test the acceptability of converting a group program into an online serious game.
Honary et al [22], 2018	Relatives Education And Coping Toolkit (RE-ACT) booklet: modular print toolkit for relatives of people with psychosis or bipolar disorder.	Converted to an online intervention with modules, resources, forums, and brief video clips. Purpose: to provide structured information and support in a web-based format.
Kingston et al [23], 2015	Paper-based maternal mental health screening tools.	Developed secure tablet and web-based e-screening for mental health, comparing perceptions of web-based vs paper-based formats. Purpose: to evaluate the acceptability and impact of web-based screening.
Lee et al [24], 2023	Promoting Independence in Dementia (PRIDE) manual-based psychosocial self-management program for mild dementia, including facilitator sessions.	Developed a web-based self-management app with similar session structure and usability enhancements (layout, login, and icons). Purpose: to support digital access and tailor the interface for users with dementia.
Monnet et al [25], 2023	Levenswensen (Life Wishes) cards supporting discussion on advance care planning in person.	Integrated into a web-based advance care planning tool designed for people with dementia and their family caregivers; developed through a user-centered iterative design process. Purpose: to create an online platform for advance care planning discussions.
Peels et al [26], 2012	Active Plus computer-tailored printed intervention to promote physical activity among adults aged ≥50 years.	The intervention was translated to a web-based format using the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework; content remained the same, but interactive outputs differed (eg, static pictures → videos, printed maps → embedded maps, added hyperlinks). Purpose: to provide tailored advice via the internet with similar core content.
Staffileno et al [27], 2015	In-person lifestyle intervention using manuals, tracking logs, and group sessions for physical activity and dietary change among young African American women.	Restructured into digital modules with multimedia components (videos, digital logs, and discussions) accessible on multiple devices. Purpose: to make behavior change content available electronically and support user self-direction.
Stille et al [28], 2015	Parent-Provider Partnerships in Referral Communication (P3RC) structured form to improve provider communication about referrals via paper or traditional processes.	Adapted to an electronic environment, integrated with digital clinical workflows and electronic health record systems. Purpose: to support electronic communication within practice systems.
Svarre et al [29], 2017	Housing Enabler paper instrument assessing environmental barriers.	Developed a digital app with item navigation, automated scoring, and data export. Purpose: to support data capture and usability in digital form.
Van Goethem et al [30], 2023	Original FOCUS program: a face-to-face psychoeducational dyadic intervention for people with advanced cancer and family caregivers.	Developed a self-managed web-based psychoeducational eHealth program (iFOCUS) using Scrum methodology for iterative development and usability testing. Purpose: to describe development and evaluate the acceptability of the web-based program.
Whitehouse et al [31], 2013	Paper-based Adolescent Screening Questionnaire for psychosocial health screening.	Converted to an eHealth platform (TickiT) cocreated with youth; adaptations included format changes for digital delivery. Purpose: to design and evaluate an online screening platform.

XSL•FO
RenderX

Strategies and Methodologies Used in the Transition Process

All 11 tools described a structured process to support the transition from paper-based to web-based formats, although the level of detail and the guiding frameworks varied (see Table 3). User-centered design was explicitly reported in 4 studies [22,24,25,30], while several additional tools applied closely related participatory or cocreation approaches [25,29,31]. This included strategies that involved users, such as patients, caregivers, and health care professionals, in interviews, workshops, think-aloud sessions, and cocreation exercises to ensure the digital tools were aligned with users' needs and practical contexts of use. These approaches involved patients, caregivers, health care professionals, or other stakeholders in workshops, interviews, think-aloud sessions, prototype reviews, and co-design activities to ensure alignment with users' needs and real-world contexts of use.

XSL•FO
RenderX

Table . Methodologies and strategies used to support the transition from paper-based to web-based social health care tools.

Study	Methodology (guiding approach or framework)[a]	Strategies (concrete methods or activities)[a]	Description
Bayley and Brown [21], 2015	Intervention mapping; theory-informed adaptation	Formative research; scenario scripting; role-play game development; message-framing survey; pre-post feasibility evaluation	Core components of the original program were reviewed for theoretical alignment and digital feasibility. Formative research informed development of an interactive role-play format
Honary et al [22], 2018	User-centered design	Workshops; prototype review; heuristic evaluation; think-aloud testing; short field testing	The web-based intervention was developed through workshops and iterative prototype evaluation, including heuristic and usability testing to inform refinements
Kingston et al [23], 2015	Randomized controlled trial as evaluation design	Digital implementation of validated tools; random assignment; secure data management	The transition to digital screening was evaluated using a randomized design comparing paper-based and electronic formats within clinical practice
Lee et al [24], 2023	User-centered design; agile development	Stakeholder consultations; wireframe reviews; think-aloud testing; iterative sprints	The app was developed through stakeholder engagement and iterative refinement, incorporating usability testing across development stages
Monnet et al [25], 2023	User-centered and participatory design; iterative agile development; Medical Research Council (MRC) framework	Cocreation; think-aloud testing; surveys and interviews; design sprints	The advance care planning tool was developed through participatory design and iterative testing with stakeholders, integrating feedback across development phases
Peels et al [26], 2012	Mixed methods pretesting and usability evaluation	Questionnaires; observations; expert and nonexpert usability testing; iterative refinement	The intervention was refined through quantitative and qualitative pretesting, including usability assessments and feedback-driven revisions
Staffileno et al [27], 2015	Phased development and feasibility evaluation	Formative sessions; digitization into web modules; usability testing; beta testing	The intervention was digitized through staged development, incorporating formative feedback and iterative usability testing prior to implementation
Stille et al [28], 2015	Practice-based implementation adaptation to electronic environments	Pilot testing; collaboration with IT teams; workflow mapping; electronic integration	The intervention was adapted for electronic clinical environments through pilot testing and integration into existing digital workflows
Svarre et al [29], 2017	Participatory research design	Observations and interviews; codesign workshops; prototype development; usability testing; iterative refinement	The digital assessment tool was developed through participatory workshops and prototype testing, incorporating iterative usability feedback
Van Goethem et al [30], 2023	User-centered design; Scrum/agile methodology	Iterative development sprints; stakeholder consultations; usability testing; feedback loops	The web-based program was developed using an iterative agile process with ongoing stakeholder input and usability testing to guide refinements
Whitehouse et al [31], 2013	Cocreation; iterative design and evaluation; mixed methods feasibility study	Stakeholder cocreation; pilot feasibility testing	TickiT was developed through iterative cocreation with youth and health care stakeholders, followed by feasibility evaluation in clinical settings

[a]"Methodologies" refer to overarching development approaches or guiding frameworks (eg, user-centered design, participatory design, Agile/Scrum, Intervention Mapping), whereas "strategies" refer to concrete activities used within these approaches (eg, cocreation workshops, interviews, think-aloud usability testing, pilot testing, and heuristic evaluation).

XSL•FO
RenderX

Iterative development processes were common. Agile or Scrum-based methodologies were described in 3 studies [24,25,30], while other tools incorporated iterative prototype testing and refinement cycles supported by usability testing and stakeholder feedback [22,29,31]. “What Should We Tell the Children?” (WSWTTC) applied Intervention Mapping to ensure theoretical alignment and structured adaptation to a game format [21]. Other studies reported phased development, mixed methods pretesting, pilot implementation, or randomized evaluation designs to assess the feasibility, usability, or effectiveness of the digital adaptation [23,26,28]. While these studies did not specify a formal design framework, they described systematic evaluation strategies to support and assess the transition process.

Challenges Encountered and Solutions Implemented

Usability was a frequently reported barrier (n=6; [22,25-27,30,31]), especially in tools aimed at older adults or people with limited digital skills. Users experienced difficulties with navigation, unclear instructions, small fonts, or confusing icons. Other tools (n=2; [21,32]) struggled with user engagement, reporting high dropout rates or limited adherence, often linked to overly long text, limited interactive elements (eg, quizzes or exercises), or cognitive burden. Technical challenges (n=4; [22,28,30,31]) included bugs, login issues, and difficulties integrating the tools into existing clinical systems. Administrative and implementation delays, such as navigating complex information and communication technology systems or securing institutional approvals, further complicated the transition process.

To address usability issues, several teams adopted iterative design strategies informed by user testing [22,26,30,31], resulting in clearer interfaces, instructional videos, customizable font sizes, and more intuitive navigation. Tools with engagement problems revised their content and structure, for example, reducing text volume, simplifying feedback messages, and adding interactive components such as videos, exercises, or tailored feedback. Technical bugs were tracked and resolved through internal reporting systems and regular meetings with development teams. In some tools [24,26,29], more pragmatic solutions, such as combining digital and print components or retaining the original structure with only minor digital adaptations, were adopted to stay within project constraints.

Discussion

Principal Findings

This scoping review examined how social health care tools have been adapted from paper-based to web-based formats. While the content of most tools remained largely unchanged, the transition often involved adding specific digital features, such as videos, online modules, quizzes, interactive exercises, adaptive navigation, and tailored feedback. User-centered and iterative development approaches were central to the transition processes, helping to identify usability barriers and refine the layout, navigation, and content presentation. At the same time, the shift to a web-based format introduced new challenges, particularly related to digital literacy, login procedures, navigation difficulties, and sustained user engagement. Most solutions to mitigate these challenges were pragmatic and context-specific, including simplifying interfaces, reducing text volume, adjusting font sizes, and combining digital components with existing workflows, highlighting the importance of adaptable and context-sensitive design when moving tools into digital environments.

This scoping review contributes to the field by focusing on the transition process from paper-based to web-based social and health care tools, rather than on effectiveness outcomes or digital tools as end products. Unlike existing reviews, it synthesizes how adaptations are made in practice and which strategies and challenges are reported. These insights can inform developers, researchers, and policymakers seeking to plan more usable, inclusive, and context-sensitive digital transitions in social and health care settings.

User-centered and iterative design approaches were frequently reported across the tools included in this scoping review, describing transitions from paper-based to web-based formats. In the included studies, these approaches were associated with reported improvements in usability, acceptability, or feasibility of the adapted tools. By involving users early and repeatedly throughout the development and usability testing process, developers were able to identify practical challenges and refine digital features accordingly. These approaches are supported by the existing literature. For instance, a recent scoping review by Evans et al [33] reported that digital health interventions developed using user-centered design methodologies were more likely to be perceived as acceptable and feasible in real-world settings, particularly among vulnerable populations. Similarly, a study by Williams and Villa [34] found that involving patients and caregivers in the cocreation process was associated with the development of digital health tools that better aligned with user needs and usability requirements.

This review also showed that even when tools transitioned from paper-based to digital formats, most retained their original content with only modest functional upgrades. Enhancements, such as incorporating multimedia elements or interactive modules, were often aimed to translate the original tool into a digital format rather than redesigning the intervention logic or user pathway. Several tools therefore focused on digitalizing delivery while keeping the core content and structure largely unchanged. At the same time, digital platforms can enable additional functionalities, such as tailored feedback, adaptive navigation, and the integration of user-entered data. Ha et al [35], for example, found that integrating patient-generated health data into digital tools was associated with improved communication and self-management in older adults. In addition, adapting paper-based interventions into digital formats can support accessibility by enabling remote access, self-paced use, and the inclusion of supportive features such as videos or structured guidance. The possibility of tailoring digital health tools to meet users’ needs, such as simplifying language and interface design, can improve engagement and may support better health-related outcomes, particularly among underserved populations [36].

This review also showed that adaptation choices were influenced by the target group. Tools for patients and family caregivers included changes related to usability and accessibility, such as

XSL•FO
RenderX

simpler navigation, adapted login systems, and supportive content. Tools for health care professionals more often focused on efficiency, data management, and integration into existing digital workflows. However, these differences were rarely described explicitly, and few studies clearly linked the characteristics of the target group to specific adaptation decisions. Similar patterns are reported in the broader digital health literature, where patient-facing tools focus more on user-centered design and professional tools focus more on workflow integration and efficiency [36]. The limited reporting on target group–specific design decisions suggests that this area needs more systematic attention in future research.

The findings of this review also demonstrate a tension between the intention to improve accessibility and the emergence of new barriers following the transition from paper-based to digital formats. Several tools in our review aimed to extend accessibility by adapting existing tools to an online format. However, this shift also created unintended challenges. For example, usability issues were especially evident in tools targeting older adults or people with limited digital skills, highlighting that increased online availability does not automatically ensure equitable access or sustained use. This mirrors the observations of Veinot et al [37], who argue that digital health innovations often privilege already-resourced users, while vulnerable groups may experience these tools as exclusionary. A systematic review by Crawford and Serhal [38] emphasized that digital transitions must be guided by a digital health equity framework, one that accounts not only for infrastructure and literacy gaps but also for users' varying emotional readiness and cognitive capacities. Considering these ethical concerns, future development should not focus solely on expanding online availability but should also incorporate inclusive design strategies and, where appropriate, complementary offline options to reduce the risk of widening existing inequities.

Lastly, an important consideration emerging from this review is the need to design digital tools that are flexible enough to function across varied digital environments and resource settings. Some tools included in our scoping review encountered implementation delays or technical issues due to rigid design or integration challenges. To support ongoing usability, digital tools should be built with adaptability in mind (eg, scalable to different systems, compatible with both high- and low-resource contexts, and structured to allow for regular updates as user needs, technologies, and care practices evolve) [39]. As Greenhalgh et al [40] highlight, sustainable digital health interventions must be both technically adaptable and socially responsive to remain relevant and usable over time.

Strengths and Limitations

This scoping review examined how social health care tools have transitioned from paper-based to web-based formats, a topic that remains relatively underexplored. A key strength of the review is its focus on describing how these transitions were carried out in practice. By analyzing the characteristics of the included tools, such as the design strategies, development approaches, and reported challenges, this study provides insights that may be useful for developers and implementers undertaking similar processes. The inclusion of both academic and gray literature (used to supplement missing information on tools already included) allowed for a more detailed description of the transition processes documented.

However, several limitations should be acknowledged. Many social health care tools may have undergone digital adaptation without formal publication, particularly those developed within routine clinical or community practice. As gray literature was not used to identify additional tools and was used only to supplement information on the included tools, relevant initiatives may not have been captured. In addition, the review was restricted to English-language sources, which may have resulted in the exclusion of tools developed in other linguistic or regional contexts. These limitations point to the need for more transparent and detailed reporting on the transition and implementation of digital health tools.

Conclusion

This scoping review highlights the importance of considering adaptability when transitioning social health care tools to digital formats, particularly in relation to different systems, resource settings, and user characteristics. Adapting tools to users' needs, for example, by simplifying navigation, adjusting content presentation, or incorporating interactive elements, may support usability and engagement. The findings suggest that user-centered, iterative, and participatory approaches are commonly used during such transitions and may assist in identifying practical challenges during development and implementation. Continued attention to reporting design decisions and contextual factors will strengthen future research on digital transitions in social and health care settings.

Acknowledgments

The authors declare the use of generative AI in the writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to generative AI tools under full human supervision: (1) editing and proofreading and (2) text refinement and shortening. The generative AI tool used was ChatGPT (GPT-5.2; OpenAI). Responsibility for the final manuscript lies entirely with the authors. Generative AI tools are not listed as authors and do not bear responsibility for the final manuscript. Declaration submitted by Charlèss Dupont.

AI was used to support the drafting of the initial version of the manuscript and to assist with revisions in response to reviewers' comments, including specific requests to reduce overly promotional language and clarify phrasing. It was also used to improve clarity, correct spelling and grammar, and shorten the text in the tables to improve readability. This support helped the authors, as they are nonnative English speakers.

XSL•FO
RenderX

Funding

This project was funded by the Swedish Research Council for Health, Working Life and Welfare (dnr: 2023-01235).

Data Availability

All data generated or analyzed during this study are included in this published paper and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string in peer-reviewed databases.
[DOCX File, 22 KB - ojphi_v18i1e88631_app1.docx]

Checklist 1
PRISMA scoping review checklist.
[PDF File, 512 KB - ojphi_v18i1e88631_app2.pdf]

References

1. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407. [doi: 10.3390/ijerph20043407] [Medline: 36834105]
2. Ogundiya O, Rahman TJ, Valnarov-Boulter I, Young TM. Looking back on digital medical education over the last 25 years and looking to the future: narrative review. J Med Internet Res 2024 Dec 19;26:e60312. [doi: 10.2196/60312] [Medline: 39700490]
3. Carini E, Villani L, Pezzullo AM, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021 Sep 8;23(9):e26189. [doi: 10.2196/26189] [Medline: 34494966]
4. Abernethy A, Adams L, Barrett M, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022;2022. [doi: 10.31478/202206e] [Medline: 36177208]
5. Lytras MD, Housawi A, Alsaywid BS, editors. Digital Transformation in Healthcare in Post-COVID-19 Times: Academic Press; 2023.
6. Tumuhimbise W, Theuring S, Atukunda EC, et al. Enablers and challenges of integrating digital health into medical education curricula: a scoping review. Discov Educ 2025 Aug 25;4(1):316. [doi: 10.1007/s44217-025-00793-z]
7. Zhou X, Smith CJM, Al-Samarraie H. Digital technology adaptation and initiatives: a systematic review of teaching and learning during COVID-19. J Comput High Educ 2023 Apr 26:1-22. [doi: 10.1007/s12528-023-09376-z] [Medline: 37359041]
8. Badr J, Motulsky A, Denis JL. Digital health technologies and inequalities: a scoping review of potential impacts and policy recommendations. Health Policy 2024 Aug;146:105122. [doi: 10.1016/j.healthpol.2024.105122] [Medline: 38986333]
9. Mumtaz H, Riaz MH, Wajid H, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023;5:1203945. [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]
10. Iyanna S, Kaur P, Ractham P, Talwar S, Najmul Islam AKM. Digital transformation of healthcare sector. What is impeding adoption and continued usage of technology-driven innovations by end-users? J Bus Res 2022 Dec;153:150-161. [doi: 10.1016/j.jbusres.2022.08.007]
11. Dong J, Jiang M, Xing Y, Zhang D, Kong J, Wang H. Challenges and experiences in digital health equity: a qualitative exploration of the implementation of the Health Code system in China. Int J Equity Health 2025 Dec 2;24(1):351. [doi: 10.1186/s12939-025-02720-y] [Medline: 41331464]
12. Xue Z, Zhou N, Wu Y, Qi H. Digital exclusion in healthcare services: a scoping review. Int J Equity Health 2026 Jan 16;25(1):37. [doi: 10.1186/s12939-025-02755-1] [Medline: 41546005]
13. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506. [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
14. Peels DA, Bolman C, Golsteijn RHJ, et al. Long-term efficacy of a printed or a web-based tailored physical activity intervention among older adults. Int J Behav Nutr Phys Act 2013 Sep 2;10(1):104. [doi: 10.1186/1479-5868-10-104] [Medline: 24228862]
15. Fontaine G, Zagury-Orly I, Maheu-Cadotte MA, et al. A meta-analysis of the effect of paper versus digital reading on reading comprehension in health professional education. Am J Pharm Educ 2021 Nov;85(10):8525. [doi: 10.5688/ajpe8525] [Medline: 34301544]

XSL•FO
RenderX

16. Pieterse M, Kip H, Cruz-Martínez RR. The complexity of eHealth implementation. In: Kip H, Beerlage-de Jong N, van Gemert-Pijnen L, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development, 1st edition: Routledge; 2018:247-270. [doi: 10.4324/9781315385907-12]
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. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
19. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis: Joanna Briggs Institute (JBI); 2020. [doi: 10.46658/JBIRM-20-01]
20. Khalil H, Campbell F, Danial K, et al. Advancing the methodology of mapping reviews: a scoping review. Res Synth Methods 2024 May;15(3):384-397. [doi: 10.1002/jrsm.1694] [Medline: 38169156]
21. Bayley JE, Brown KE. Translating group programmes into online formats: establishing the acceptability of a parents' sex and relationships communication serious game. BMC Public Health 2015 Dec 9;15(1):1225. [doi: 10.1186/s12889-015-2545-0] [Medline: 26651616]
22. Honary M, Fisher NR, McNaney R, Lobban F. A web-based intervention for relatives of people experiencing psychosis or bipolar disorder: design study using a user-centered approach. JMIR Ment Health 2018 Dec 7;5(4):e11473. [doi: 10.2196/11473] [Medline: 30530457]
23. Kingston D, Janes-Kelley S, Tyrrell J, et al. An integrated web-based mental health intervention of assessment-referral-care to reduce stress, anxiety, and depression in hospitalized pregnant women with medically high-risk pregnancies: a feasibility study protocol of hospital-based implementation. JMIR Res Protoc 2015 Jan 16;4(1):e9. [doi: 10.2196/resprot.4037] [Medline: 25595167]
24. Lee AR, Csipke E, Yates L, et al. A web-based self-management app for living well with dementia: user-centered development study. JMIR Hum Factors 2023 Feb 24;10:e40785. [doi: 10.2196/40785] [Medline: 36826978]
25. Monnet F, Dupont C, Smets T, et al. Advance care planning website for people with dementia and their family caregivers: protocol for a development and usability study. JMIR Res Protoc 2023 Jul 26;12:e46935. [doi: 10.2196/46935] [Medline: 37494084]
26. Peels DA, Bolman C, Golsteijn RHJ, et al. Differences in reach and attrition between web-based and print-delivered tailored interventions among adults over 50 years of age: clustered randomized trial. J Med Internet Res 2012 Dec 17;14(6):e179. [doi: 10.2196/jmir.2229] [Medline: 23246790]
27. Staffileno BA, Tangney CC, Fogg L, Darmoc R. Making behavior change interventions available to young African American women: development and feasibility of an eHealth lifestyle program. J Cardiovasc Nurs 2015;30(6):497-505. [doi: 10.1097/JCN.0000000000000197] [Medline: 25325372]
28. Stille CJ, Lockhart SA, Maertens JA, Madden CA, Darden PM. Adapting practice-based intervention research to electronic environments: opportunities and complexities at two institutions. EGEMS (Wash DC) 2015;3(1):1111. [doi: 10.13063/2327-9214.1111] [Medline: 25848633]
29. Svarre T, Lunn TBK, Helle T. Transforming paper-based assessment forms to a digital format: exemplified by the Housing Enabler prototype app. Scand J Occup Ther 2017 Nov;24(6):438-447. [doi: 10.1080/11038128.2016.1255774] [Medline: 27842449]
30. Van Goethem V, Dierickx S, Matthys O, et al. A self-management psychoeducational eHealth program to support and empower people with advanced cancer and their family caregivers: development using the scrum methodology. Internet Interv 2023 Sep;33:100659. [doi: 10.1016/j.invent.2023.100659] [Medline: 37593143]
31. Whitehouse SR, Lam PY, Balka E, et al. Co-creation with TickiT: designing and evaluating a clinical eHealth platform for youth. JMIR Res Protoc 2013 Oct 18;2(2):e42. [doi: 10.2196/resprot.2865] [Medline: 24140595]
32. Boots LM, de Vugt ME, Withagen HE, Kempen GI, Verhey FR. Development and initial evaluation of the web-based self-management program "Partner in Balance" for family caregivers of people with early stage dementia: an exploratory mixed-methods study. JMIR Res Protoc 2016 Mar 1;5(1):e33. [doi: 10.2196/resprot.5142] [Medline: 26932438]
33. Evans L, Evans J, Pagliari C, Källander K. Scoping review: exploring the equity impact of current digital health design practices. Oxf Open Digit Health 2023 Jan 1;1:qad006. [doi: 10.1093/oodh/oqad006]
34. Williams J, Villa A. Co-creating a digital first integrated care experience with patients and caregivers. Int J Integr Care 2025 Apr 9;25:587. [doi: 10.5334/ijic.ICIC24275]
35. Ha S, Ho SH, Bae YH, et al. Digital health equity and tailored health care service for people with disability: user-centered design and usability study. J Med Internet Res 2023 Nov 28;25:e50029. [doi: 10.2196/50029] [Medline: 38015589]
36. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and facilitators of digital health among culturally and linguistically diverse populations: qualitative systematic review. J Med Internet Res 2023 Feb 28;25:e42719. [doi: 10.2196/42719] [Medline: 36853742]
37. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
38. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 2;22(6):e19361. [doi: 10.2196/19361] [Medline: 32452816]

XSL•FO
RenderX

39. Maqbool B, Herold S. Potential effectiveness and efficiency issues in usability evaluation within digital health: a systematic literature review. J Syst Softw 2024 Feb;208:111881. [doi: 10.1016/j.jss.2023.111881]
40. Greenhalgh T, Wherton J, Papoutsi 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 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
WSWTTC: What Should We Tell the Children?

Edited by E Mensah; submitted 28.Nov.2025; peer-reviewed by T Svarre, WKH Mok; accepted 11.Jul.2026; published 06.Aug.2026.

Please cite as:
Dupont C, Savolainen E, Fjällström P, Eneslätt M
Adaptation of Social Health Care Tools From Paper-Based to Web-Based Formats: Scoping Review
Online J Public Health Inform 2026;18:e88631
URL: https://ojphi.jmir.org/2026/1/e88631
doi:10.2196/88631

© Charlèss Dupont, Eva Savolainen, Petter Fjällström, Malin Eneslätt. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 6.Aug.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Safety-Oriented Evaluation of Large Language Models in Health Care: Guideline-Informed Systematic Review

Hikaru Matsuoka[1,2,3*], PhD; Takayuki Takahashi[3,4*], MD, PhD; Takayuki Semitsu[2*], MSc; Keiko Matsuoka[5*], MD; Ryoma Hayashi[4*], MD

[1]Department of Information Security, Faculty of Information Systems, University of Nagasaki, Nagasaki, Nagasaki, Japan

[2]Information-Technology Promotion Agency, Minato, Tokyo, Japan

[3]Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Nihonbashi 1-chome Mitsui Building 15F, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan

[4]Department of Obstetrics and Gynecology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan

[5]Department of Obstetrics and Gynecology, Shonan Atsugi Hospital, Atsugi, Kanagawa, Japan

[*]all authors contributed equally

Corresponding Author:
Takayuki Takahashi, MD, PhD
Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Nihonbashi 1-chome Mitsui Building 15F, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan

Abstract

Background: Large language models (LLMs) are rapidly emerging in health care, offering opportunities in decision support, education, and research, but raising critical concerns about safety, reliability, and ethics. Although several guidelines for trustworthy AI exist in business and technology, few systematic reviews have applied them to medical contexts.

Objective: This study aimed to conduct a systematic review of LLM research in health care, applying the AI Guidelines for Business as a framework across 11 domains, including safety, reliability, ethics, transparency, fairness, inclusiveness, privacy, security, robustness, data quality, and verifiability.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (retrospectively registered on the Open Science Framework; DOI 10.17605/OSF.IO/P4KSB), the PubMed, Scopus, Web of Science, arXiv, and IEEE Xplore databases were searched on January 15, 2025. Records were screened in 2 stages by 3 reviewers (with records retained only upon unanimous agreement). A total of 247 studies were included, of which 211 (85.4%) contributed quantitative values. Eligible studies were classified across 11 trustworthy AI domains. Heterogeneous metrics were summarized within metric families; when multiple models were evaluated, the mean across models was used as the primary estimate, with best, median, and primary-model sensitivity analyses. The LLM-assisted categorization (GPT-5 mini) was validated by using an automated internal consistency check, and 95% CIs were estimated by using cluster bootstrap on study-level values.

Results: Of the 25,156 records, 247 (1.0%) studies were included, and of these, 211 (85.4%) contributed quantitative values. Evaluation concentrated on accuracy (143/247, 57.9%) and fairness and inclusiveness (93/247, 37.7%), followed by data quality (47/247, 19.0%) and prevention of misinformation (44/247, 17.8%). Normalized performance was moderate to high (accuracy mean 0.73, 95% CI 0.7-0.76; data quality: 0.64; prevention of misinformation: 0.82). Selecting the best-performing model inflated domain means by up to 0.05. Privacy protection (2/247, 0.8%) and security assurance (0/247, 0.0%) were almost entirely absent. Domain assignments were recoverable from objective metric types in 98.9% of values (Cohen κ=0.985).

Conclusions: Current evaluations emphasize accuracy while underreporting privacy, security, robustness, explainability, and verifiability. This finding reflects gaps in reporting rather than demonstrated poor performance, underscoring the need for comprehensive, guideline-based, multidomain evaluation before deployment in high-stakes clinical settings.

(Online J Public Health Inform 2026;18:e97240) doi:10.2196/97240

KEYWORDS

artificial intelligence; AI; large language model; patient safety; systematic review; trustworthy artificial intelligence; trustworthy AI; health care informatics; clinical decision support; guideline; evaluation framework; public health informatics

XSL•FO
RenderX

Introduction

The adoption of large language models (LLMs) within health care has accelerated sharply, fueled by advances in generative AI and the growing demand for tools to support diagnostics, clinical documentation, patient communication, medical education, and clinical decision-making [1-4]. The potential of these models to synthesize vast amounts of medical knowledge and streamline clinical workflows is significant [5,6]. Yet, alongside this promise, concerns over their safety, reliability, ethical implications, fairness, transparency, robustness, accountability, privacy, and overall trustworthiness have become increasingly salient [7-10]. Although many early studies focused on assessing the diagnostic accuracy or utility of LLMs, particularly on standardized examinations [11,12], fewer have systematically evaluated whether these models conform to comprehensive guideline-based standards for responsible AI or whether domains such as safety and ethical risk are adequately addressed in practice [13,14].

Recent systematic reviews have begun to shed light on these multifaceted issues. For instance, assessments of LLM performance under examination conditions reveal moderate to high accuracy but also significant variability across studies and medical specialties [15,16]. Other research highlights the critical risks to patient safety and the increased clinician burden that can arise from the uncritical deployment of generative AI, emphasizing the need for robust harm reduction strategies [17,18]. Studies exploring user perceptions in specialized fields such as mental health underscore the importance of safety guardrails and technical controls to mitigate potential harms [19]. Furthermore, critical assessments of the use of LLMs in research, such as their use in systematic reviews, have identified substantial methodological challenges and a high risk of factual errors or hallucinations, which could compromise the integrity of evidence synthesis [20,21]. Together, these studies demonstrate that although performance benchmarks are steadily advancing, structured and comprehensive safety assessment remains a critical but underdeveloped area [13].

Concurrently, a robust ecosystem of guideline frameworks has emerged from academic, industry, and international initiatives to define the principles of trustworthy and responsible AI. Reporting guidelines such as CONSORT (Consolidated Standards of Reporting Trials)-AI and SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials)-AI provide frameworks for standardizing the evaluation of clinical trials involving AI interventions, promoting transparency and reproducibility [22,23]. The FUTURE (Fairness, Universality, Traceability, Usability, Robustness, and Explainability)-AI guideline offers an international consensus on 6 core principles—fairness, universality, traceability, usability, robustness, and explainability, as well as best practices for the entire lifecycle of medical AI systems, particularly in medical imaging [24]. Broader mapping studies, including a review of 200 guidelines and recommendations for AI governance, have identified recurring ethical principles across international frameworks [25]. In Japan, the AI Guidelines for Business, jointly compiled by the Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry, similarly highlight key areas for responsible implementation, including safety, reliability, ethics, transparency, and data quality [26].

This development of best practices is mirrored by growing regulatory momentum worldwide. In the European Union, the AI Act is set to impose stringent, risk-based requirements on high-risk AI systems—a category that includes many health care applications—mandating rigorous risk assessment, transparency, conformity, and postmarket monitoring [27]. In the United States, the Food and Drug Administration (FDA) is actively advancing its regulatory framework for AI- or machine learning (ML)–based medical devices, emphasizing the need for a total product lifecycle approach to ensure safety and effectiveness [28]. At the global level, the World Health Organization (WHO) has published landmark guidance emphasizing that AI systems for health must be designed to respect human rights, mitigate bias, ensure safety and privacy, and maintain clear lines of accountability [29]. These convergent developments reflect a clear global expectation that medical LLMs must not only perform technical tasks effectively but also adhere to a stringent set of ethical, safety, regulatory, and trustworthiness criteria.

We adopted the AI Guidelines for Business as the organizing framework for 3 reasons. First, unlike most health-specific reporting standards, which target specific study types (eg, CONSORT-AI and SPIRIT-AI for trials, TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis)+AI for prediction models, and CLAIM (Checklist for Artificial Intelligence in Medical Imaging) and STARD (Standards for Reporting Diagnostic Accuracy Studies)-AI for diagnostic and imaging studies), the framework provides a single, cross-sectoral taxonomy spanning the full range of trustworthy AI principles, allowing heterogeneous LLM evaluation studies to be mapped onto 1 consistent scheme. Second, its principle set aligns closely with international health guidance, including WHO guidance and FUTURE-AI; we make this alignment explicit in a correspondence table (Multimedia Appendix 1). Third, it is implementation and lifecycle oriented, making it well suited to a review concerned with deployment readiness. We use it as a complementary organizing scaffold rather than a substitute for health-specific reporting guidelines, against which we cross-map every domain.

This study aims to fill these gaps by conducting a systematic review of the medical LLM literature, using a comprehensive, business-derived AI guidelines framework. We will classify existing studies across multiple domains, including safety, reliability, ethics, transparency, fairness, privacy, accountability, explainability, robustness, data quality, inclusiveness, and verifiability, using structured prompts for data extraction. Furthermore, we will produce descriptive visualizations to illustrate how frequently each domain is addressed, thereby highlighting areas of considerable risk or deficiency and delineating critical research gaps. The findings aim to inform safer, more trustworthy medical LLM development, regulation, and clinical deployment. For contextual comparison, we incorporate the review methodology of Shool et al [16], focusing on the extraction and assessment of numerical data to ensure that our synthesis remains focused on objective evidence while

XSL•FO
RenderX

aligning with prior systematic efforts, thereby transparently capturing the limitations of current evaluation metrics.

Methods

Systematic Review Data Extraction and Management Protocol

The literature was systematically searched on January 15, 2025, using the PubMed, Scopus, Web of Science, arXiv, and IEEE Xplore databases. Queries combined relevant keywords and MeSH terminology such as evaluation, LLMs, AI chatbot, and medical and clinical practice. The complete search schema is presented in Multimedia Appendix 2; the title-field searches yielded 1201 (4.8%) records in PubMed, 3273 (13.0%) records in Scopus, 1989 (7.9%) records in Web of Science, 18,128 (72.1%) records in arXiv, and 565 (2.2%) records in IEEE Xplore (N=25,156 records). This review followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 (completed checklist in Checklist 1) and was retrospectively registered on the Open Science Framework (OSF; DOI 10.17605/OSF.IO/P4KSB; registered on June 22, 2026; labeled retrospective), as it was not prospectively registered. Eligible studies were original research that empirically evaluated ≥1 LLM in a health care context and reported at least 1 quantitative outcome amenable to scoring, published in English; nonoriginal work, records with inaccessible text or a missing DOI, duplicate records, and studies without original quantitative evaluation were excluded. Preprints were identified by their archive URL (arXiv or medRxiv) and examined in a sensitivity analysis. Records were screened in 2 stages (title and abstract, followed by full-text screening) by 3 reviewers working together; records were retained only when all 3 reviewers agreed (ie, by unanimous consensus).

A comprehensive search across the PubMed, Scopus, Web of Science, arXiv, and IEEE Xplore databases yielded 25,156 records, from which 2754 (10.9%) duplicate records and 328 (1.3%) additional records were removed. This resulted in 22,074 (87.7%) records being screened by title and abstract, leading to the exclusion of 20,198 (91.5%) records that did not meet the inclusion criteria. Following this, data extraction was performed on 1876 (8.5%) articles that passed the initial screening. Of these, 586 (31.2%) articles were excluded because of inaccessible abstracts or full texts, missing DOIs, duplication, and nonoriginal research designs.

Following a detailed full-text review, an additional 529 (41.0%) articles were excluded. Ultimately, this rigorous and meticulous effort culminated in 761 (59.0%) articles from which data were fully extracted. After further excluding studies without original quantitative evaluation or for which reliable scoring was not feasible, a final dataset of 247 (32.5%) articles was retained for analysis.

Approach to Score Calculation and Normalization

We summarized performance within metric families (percentages, 0-1 proportions, Likert or ordinal, readability, and agreement) and did not pool fundamentally different families; metric-family-specific summaries are provided in Multimedia Appendix 3. For an auxiliary descriptive comparison only, extracted values were normalized to a 0-1 scale. Performance measures such as accuracy, the area under the curve (AUC), and the F_1-score were regarded as equivalent indicators of correctness and were normalized accordingly. For Likert-type scales (eg, 5-point scales), scores were rescaled such that the highest value (eg, 5) corresponded to the most favorable outcome. For readability indices, including the Simple Measure of Gobbledygook (SMOG) index and the Flesch-Kincaid Grade Level, the maximum level of 21, representing graduate-level complexity, was provisionally defined as the upper bound, and scores were aggregated relative to this reference point.

When a study evaluated multiple models, we applied a prespecified extraction rule and computed a per-study representative value as the mean across models (the primary estimate), with the best, median, and primary (first-reported) model used as sensitivity analyses (Multimedia Appendix 3). The analysis was descriptive, not a formal meta-analysis; to avoid pseudoreplication, we aggregated data into 1 value per study per domain and estimated 95% CIs by using cluster bootstrap resampling of studies (with categories containing <5 valid items excluded from interval estimation). For the branch-of-medicine subgroup analyses, which are based on the best-model value rather than the mean across models, 95% CIs were calculated as mean ± 1.96 × SD/√n using a normal approximation; these intervals may therefore extend beyond the 0-1 range of the normalized score.

Data Classification and Categorization

The number of publications increased between 2019 and 2025. Of the 761 articles from which data were fully extracted, 1 (0.1%) was published in 2019, followed by 3 (0.4%) in 2020, 6 (0.8%) in 2021, 7 (0.9%) in 2022, and 160 (21.0%) in 2023. The largest number was published in 2024, with 557 (73.2%) articles, followed by 27 (3.5%) in early 2025. These are descriptive publication-year counts of the 761 articles; no statistical test for change over time was performed.

In the process of study selection and data extraction, each article was analyzed to determine the corresponding medical specialty (allowing for multiple assignments), 3 independent binary indicators (emergency conditions, surgical diagnosis, and medical [internal medicine] diagnosis), which are not mutually exclusive (ie, a study may be flagged for >1), as well as the intended purpose of use and the study population. An LLM (GPT-5 mini [gpt-5-mini, OpenAI]) was used as an auxiliary tool for text categorization. We note the timeline explicitly: the literature search was conducted on January 15, 2025, whereas LLM-assisted extraction and categorization were performed in September 2025, after the public release of the GPT-5 model family on August 7, 2025; the model (GPT-5 mini), access date (September 2025), default decoding temperature, and full prompts are reported in Multimedia Appendix 4. Terms identified in each study were mapped to the categories defined in the AI Guidelines for Business and the Guide to Evaluation Perspectives on AI Safety by using the coding manual in Multimedia Appendix 5. To validate the categorization without additional human raters, we used an automated internal consistency check: domain assignments matched the metric-derived domain in 98.9% of 3646 extracted values

XSL•FO
RenderX

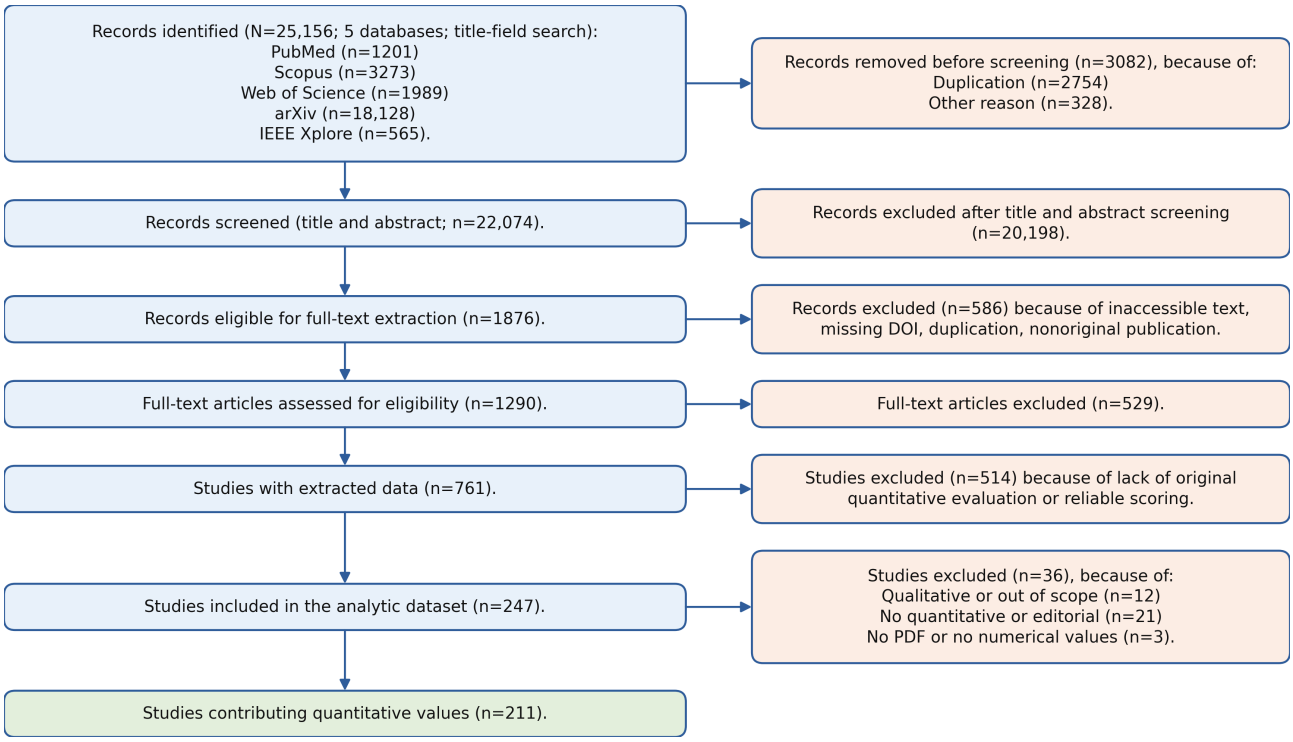
(Cohen κ=0.985; 99.0% of 1454 distinct metric types mapped to a single domain), indicating that the categorization was systematic and reproducible. Because conventional risk-of-bias tools (eg, PROBAST [Prediction Model Risk of Bias Assessment Tool]-AI) require subjective expert judgment that does not scale across this large number of cross-specialty studies, we instead derived an objective Reporting Transparency Index (RTI) from the extracted data (model named, version or date specified, sample size reported, ≥2 models compared, and ≥2 distinct metrics reported; score range 0-5) and used it in a sensitivity analysis restricted to higher-transparency studies (Multimedia Appendix 6).

Results

Study Selection

The study selection process is summarized in Figure 1. Of the 25,156 records identified across the 5 databases, 247 studies met the inclusion criteria, of which 211 contributed extractable quantitative values.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram showing study identification, screening, and inclusion. Of 25,156 records across 5 databases (title-field searches), 247 studies were included, of which 211 contributed extractable quantitative values.

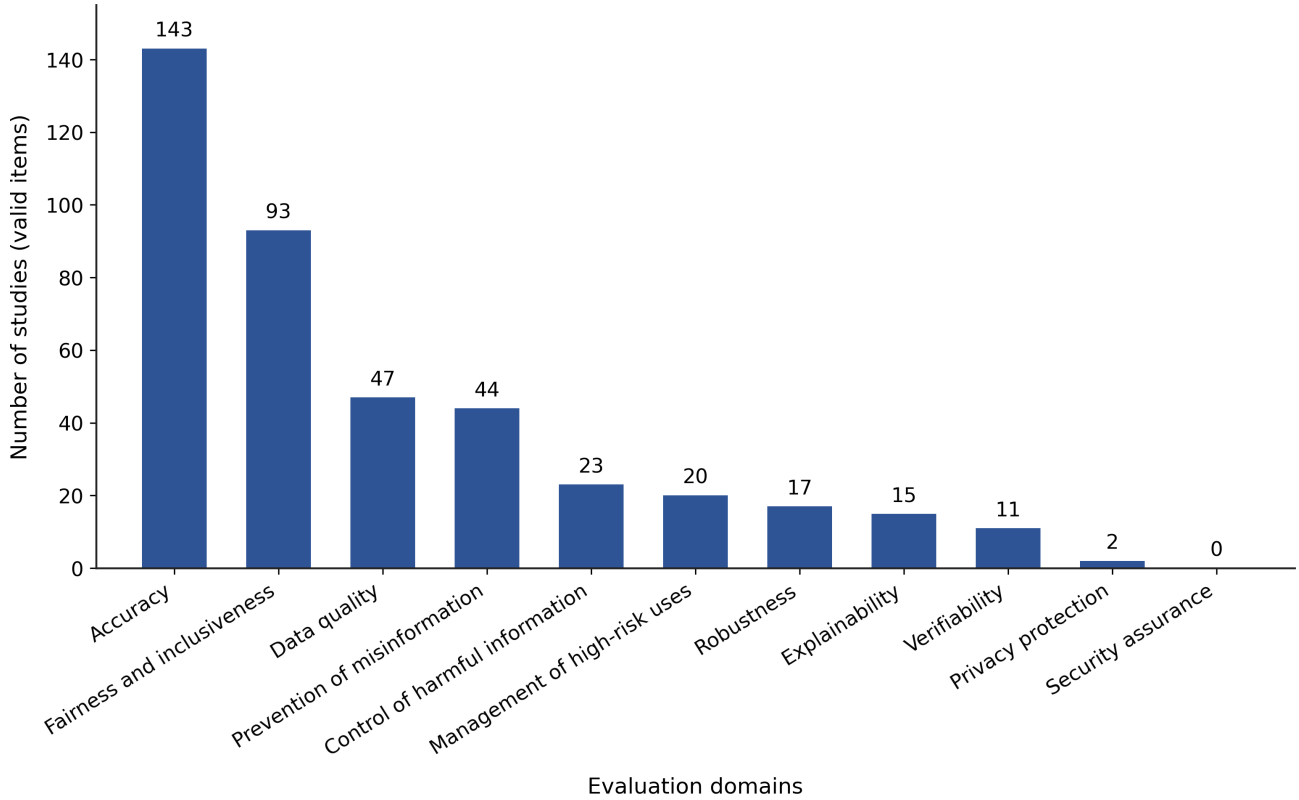


Distribution of Categories

Figure 2 presents the number of papers across multiple categories. A valid item denotes a study-domain pair with an extractable quantitative value; of the 247 included studies, 211 (85.4%) contributed valid items. This highlights which categories were more frequently addressed and which were largely overlooked.

XSL•FO
RenderX

Figure 2. Number of studies addressing each trustworthy AI evaluation domain (N=247). Accuracy was the most frequently assessed domain; privacy protection and security assurance domains were almost absent.



On the basis of the terminology exemplified in Table 1, the reported scores for each category were extracted as assessed by the authors. Quantitative metrics such as accuracy, AUC, and the F_1-score provide objective indicators of model performance and allow consistent comparison across studies. Their use facilitates transparency and reliability in evaluating LLMs for health care applications.

Table . Category-keyword correspondence used to map extracted terms to trustworthy AI domains (full coding manual in Multimedia Appendix 5).

Trustworthy AI domains	Representative terms or keywords used for mapping
Safety (overall)	safety
Control of harmful information	harm
Prevention of misinformation	misinformation; misuse; usefulness; relevance; reliability
Fairness and inclusiveness	bias; fair; acceptability; concise; comprehensive
Management of high-risk uses	risk
Privacy protection	privacy
Security assurance	security
Explainability	readability; transparency; visibility; understandability; clarity
Robustness	robust; consistency; coherence; redundancy
Data quality	quality; completeness
Verifiability	verification; actionability; reference

Accuracy is the most represented category, with 143 valid items, showing that performance measurement is the most commonly reported aspect. Fairness and inclusiveness (93 valid items), data quality (47 valid items), and prevention of misinformation (44 valid items) follow as moderately covered categories. Categories such as control of harmful information (23 valid items), management of high-risk uses (20 valid items), robustness (17 valid items), explainability (15 valid items), and verifiability (11 valid items) have relatively limited coverage, indicating less consistent reporting. Privacy protection (2 valid items) and security assurance (0 valid items) are almost entirely absent, suggesting that these categories have received comparatively little attention within the field of health care.

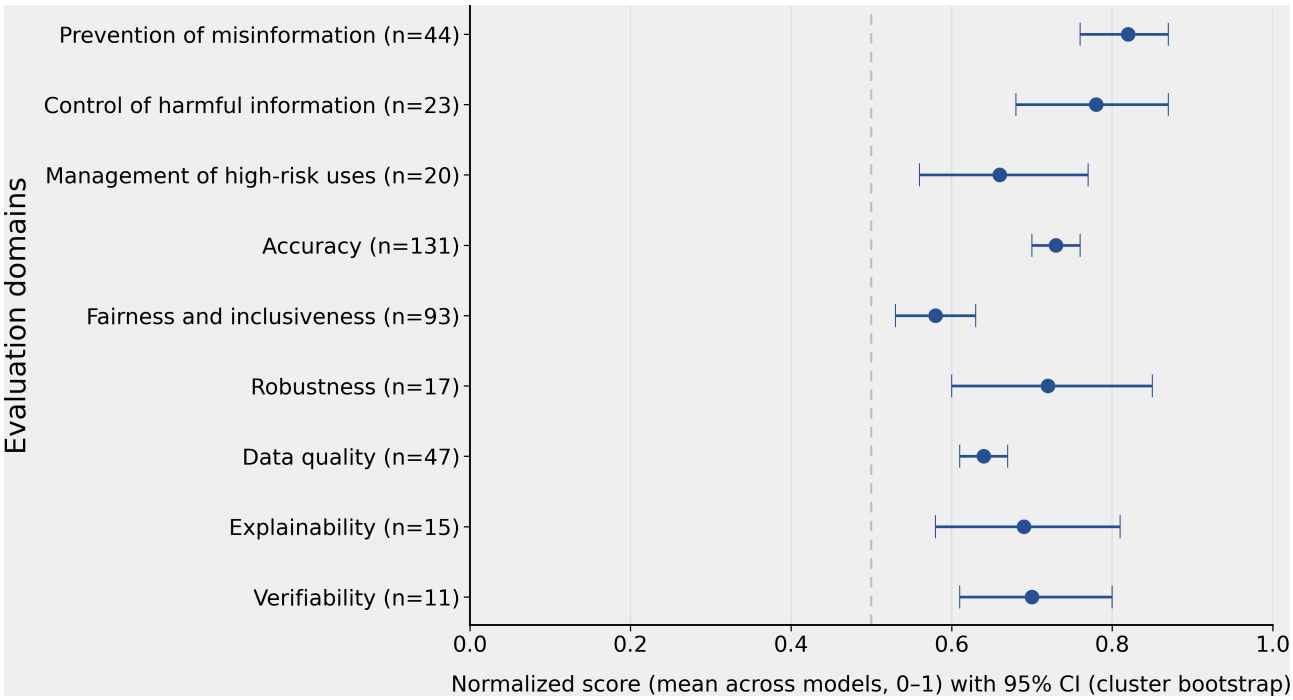
Category-Wise Numerical Evaluation

A descriptive interval plot (Figure 3) presents study-aggregated mean scores and 95% CIs (cluster bootstrap) for each evaluation category. Categories with <5 valid items were excluded to avoid

XSL•FO
RenderX

unstable estimates. Reporting metric families separately revealed substantial within-domain heterogeneity (eg, the normalized median for accuracy differed by family: Likert or ordinal, 0.85; percentage, 0.72; 0-1 proportion, 0.76; Multimedia Appendix 3).

Figure 3. Domain-level performance showing study-aggregated mean normalized score (0 - 1; mean across models) with 95% CIs estimated by cluster bootstrap. Domains with <5 valid items (privacy and security) are not shown.



Accuracy (131/247, 53.0%; mean 0.73, 95% CI 0.7-0.76) had the largest sample size and a narrow CI. Selecting the best-performing model rather than averaging inflated domain means by 0.01 to 0.05 (eg, accuracy best 0.78 vs mean 0.73); therefore, we report mean-based estimates as the primary analysis (Multimedia Appendix 3). Restricting the synthesis to higher-transparency studies (RTI≥3; 186/242) changed domain estimates by at most 0.004, and excluding the 17 preprints in the analytic set changed estimates by at most 0.018, indicating robustness. Control of harmful information (23/247, 9.3%), prevention of misinformation (44/247, 17.8%), and data quality (47/247, 19.0%) demonstrated relatively high mean scores (0.64-0.82) with moderate precision. Fairness and inclusiveness (93/247, 37.7%; mean 0.58, 95% CI 0.53-0.63) showed a lower average score but maintained a reasonably narrow CI because of the larger sample size.

Management of high-risk uses (20/247, 8.1%; mean 0.66, 95% CI 0.56-0.77), explainability (15/247, 6.1%; mean 0.69, 95% CI 0.58-0.81), robustness (17/247, 6.9%; mean 0.72, 95% CI 0.6-0.85), and verifiability (11/247, 4.5%; mean 0.7, 95% CI 0.61-0.8) all exhibited wide CIs. The wide intervals reflect considerable variability and reduced certainty of the estimates. This is mainly attributable to the small number of valid items in these categories, which made the results more sensitive to outliers and sampling error.

In summary, categories supported by larger sample sizes (such as accuracy) yielded stable estimates with narrow CIs, while categories with limited representation (explainability, robustness, verifiability, and management of high-risk uses) produced much wider CIs, highlighting substantial variability and uncertainty.

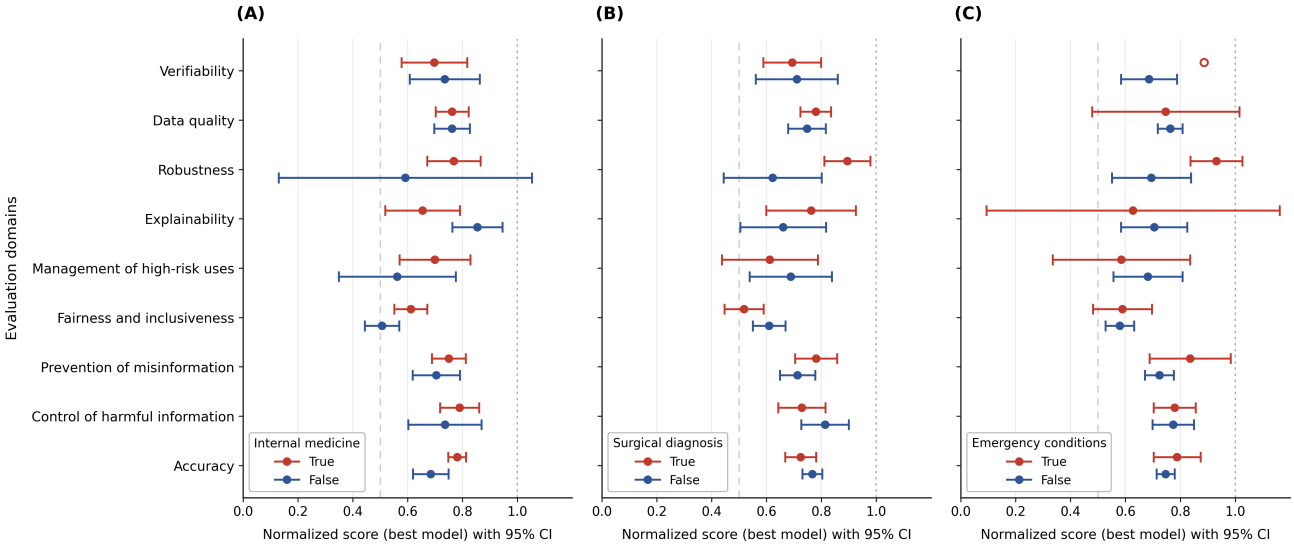
Evaluations Within Each Branch of Medicine

Internal Medicine

Here, the True and False subgroup labels denote whether each independent binary indicator (emergency, surgical, or medical diagnosis) applied to a study. In the internal medicine group (Figure 4A), accuracy was consistently high, particularly for the True subgroup (98/140, 70.0%; mean 0.78, 95% CI 0.75-0.81), although slightly lower for the False subgroup (42/140, 30.0%; mean 0.68). Control of harmful information and prevention of misinformation also showed moderate to high scores (means ~0.75), but the estimates were based on relatively smaller sample sizes and therefore displayed wider CIs. Notably, fairness and inclusiveness achieved relatively higher scores in this domain, especially in the True subgroup (66/93, 71.0%; mean 0.61, 95% CI 0.55-0.67), which stands out compared with other categories and indicates a greater emphasis on equitable evaluation within internal medicine. Data quality was another strong category, with both True and False subgroups showing mean values of approximately 0.76 and narrow CIs; small-sample categories are interpreted with caution and not described as robust. Categories such as robustness (mean 0.77) and verifiability (mean 0.7) were also represented, although the smaller number of cases led to broader variability. By contrast, management of high-risk uses and explainability were less well supported, with only limited data and correspondingly wide CIs.

XSL•FO
RenderX

Figure 4. Performance by branch-of-medicine subgroup (true vs false) for (A) internal medicine, (B) surgical diagnosis, and (C) emergency conditions. Points represent subgroup means (best-model value) with 95% CIs.



Surgical Diagnosis

In the surgical diagnosis group, accuracy remained high in both subgroups (True: 50/140, 35.7%; mean 0.72 and False: 90/140, 64.3%; mean 0.77), with relatively narrow CIs. Control of harmful information and prevention of misinformation also yielded favorable means of around 0.73 to 0.81, although they were based on modest sample sizes. Fairness and inclusiveness showed moderate values, with the False subgroup (64/93, 68.8%; mean 0.61, 95% CI 0.55-0.67) slightly higher than the True subgroup (29/93, 31.2%; mean 0.52). Data quality was consistently strong across both groups (means ~0.75-0.78). Robustness presented mixed results: although the True subgroup reached a notably high mean of 0.89, this estimate was derived from only 6 cases and thus carried wide uncertainty, whereas the False subgroup mean was lower (0.62). Other categories such as management of high-risk uses, explainability, and verifiability were represented by <15 cases and exhibited wider intervals, reflecting reduced reliability.

Overall, the interval plot (Figure 4B) demonstrates that surgical diagnosis evaluations emphasize accuracy and data quality, with moderate attention to fairness. However, caution is warranted when interpreting categories with very limited sample sizes, for which variability remains substantial.

Emergency Conditions

In the emergency conditions group, accuracy remained high across both subgroups, with the True group (17/140, 12.1%; mean 0.79; 95% CI 0.7-0.87) showing slightly higher values than the False group (123/140, 87.9%; mean 0.75; 95% CI 0.71-0.78). Control of harmful information (18/22, 81.8%; mean 0.77) and prevention of misinformation (38/43, 88.4%; mean 0.72) demonstrated moderate to strong performance, although sample sizes were relatively limited for the former. Fairness and inclusiveness showed balanced values across groups (True: 0.59, False: 0.58), suggesting consistent but moderate attention to equity. Data quality (43/46, 93.5%; mean 0.76, 95% CI 0.72-0.81) again emerged as a reliable category, with narrower CIs than those of other measures. By contrast, categories such as management of high-risk uses, explainability, robustness, and verifiability were represented by <15 cases, leading to wider intervals and greater uncertainty regarding their interpretation.

Overall, the interval plot (Figure 4C) indicates that emergency condition evaluations prioritize accuracy and data quality, while fairness remains moderate. Importantly, the evaluation of LLM performance in emergency contexts is particularly meaningful, as these scenarios often involve time-critical decisions for which reliability and safety are paramount.

Application and Study Population

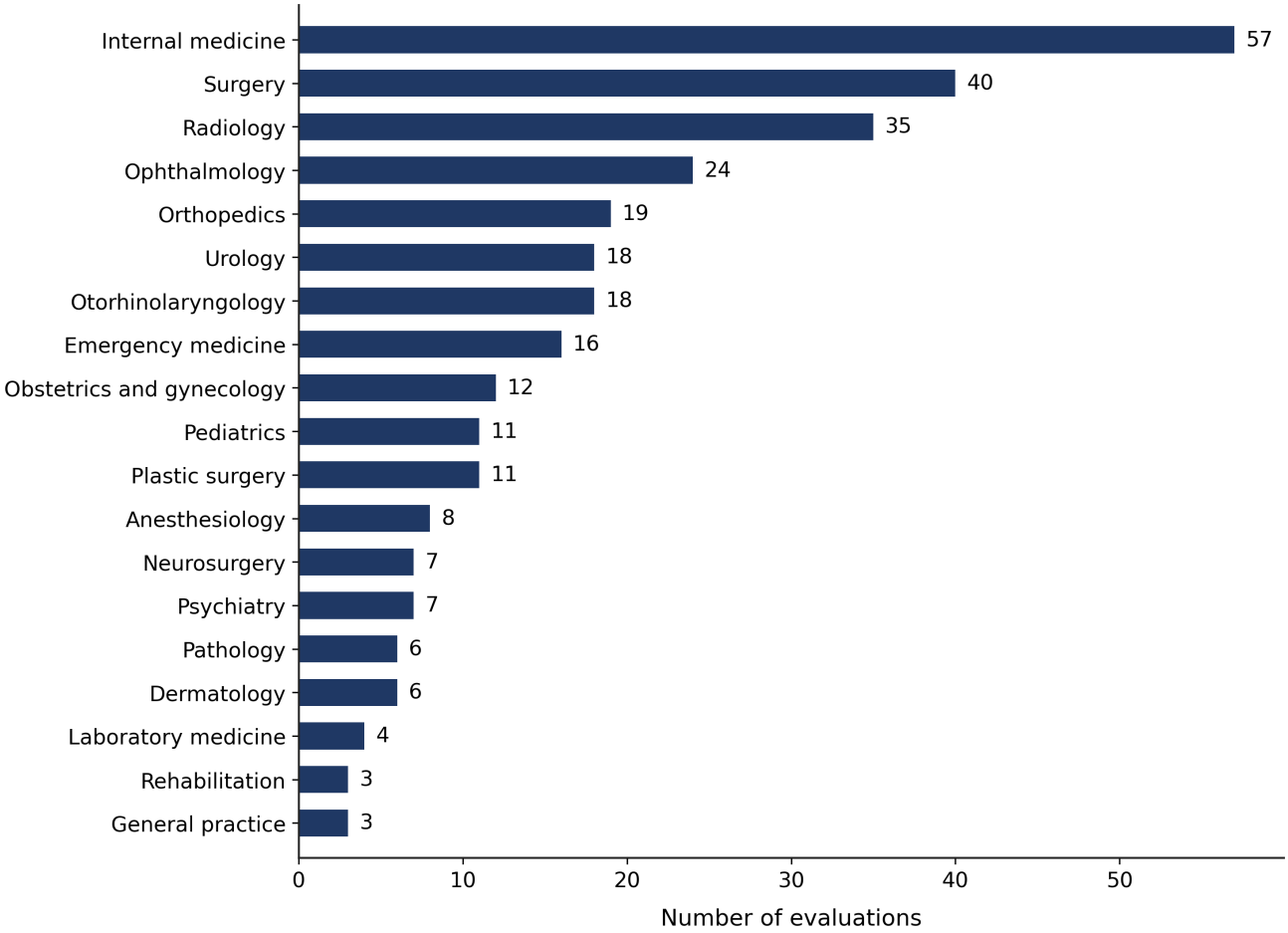
The 19 predefined medical specialties were categorized using an LLM according to the structured prompt outlined in Table 2. The distribution of evaluations across the 19 medical specialties (Figure 5) revealed substantial variation in research focus. Internal medicine accounted for the largest share (57/247, 23.1%), highlighting its central role in the assessment of LLMs. This was followed by surgery (40/247, 16.2%) and radiology (35/247, 14.2%), indicating strong representation in procedure-oriented and imaging-based fields. Ophthalmology (24/247, 9.7%), orthopedics (19/247, 7.7%), and urology (18/247, 7.3%) also contributed considerably, alongside otorhinolaryngology (18/247, 7.3%) and emergency medicine (16/247, 6.5%).

XSL•FO
RenderX

Table . Structured data extraction schema applied with a large language model (LLM)–assisted prompt (full prompt in Multimedia Appendix 4).

Extraction fields	Description or allowed values
Study population	Free text (eg, adult patients with type 2 diabetes; emergency physicians)
Medical specialty (multiple allowed)	≥1 of the following 19 specialties: internal medicine; pediatrics; dermatology; psychiatry; surgery; orthopedics; obstetrics and gynecology; ophthalmology; otorhinolaryngology; urology; neurosurgery; radiology; anesthesiology; pathology; laboratory medicine; emergency medicine; plastic surgery; rehabilitation; general practice
Emergency conditions	True or false (independent indicator)
Surgical diagnosis	True or false (independent indicator)
Medical (internal medicine) diagnosis	True or false (independent indicator)
Application	Free text (intended use; eg, clinical decision support)
LLMs evaluated	Model names and version
Reported metrics and values	For example, accuracy, area under the curve, F_1-score, Likert rating, readability index
Trustworthy AI domains	Mapped according to Table 1 (multiple domains allowed)

Figure 5. Distribution of evaluations across 19 predefined medical specialties.



Midrange representation was observed for obstetrics and gynecology (12/247, 4.9%), pediatrics (11/247, 4.5%), and plastic surgery (11/247, 4.5%), reflecting moderate engagement. Several specialties showed lower counts, including anesthesiology (8/247, 3.2%), neurosurgery (7/247, 2.8%), and psychiatry (7/247, 2.8%). At the lowest end, pathology (6/247, 2.4%), dermatology (6/247, 2.4%), laboratory medicine (4/247, 1.6%), rehabilitation (3/247, 1.2%), and general practice (3/247, 1.2%) were minimally represented.

Overall, the graph demonstrates that LLM-related evaluations have concentrated heavily in internal medicine, surgery, and radiology, while primary care and supportive specialties remain underexplored. This imbalance underscores the need for broader

XSL•FO
RenderX

coverage to ensure that AI applications are evaluated across the full spectrum of medical practice.

Discussion

Principal Results

This systematic review reveals a marked imbalance in the evaluation of medical LLMs. Our findings demonstrate that the vast majority of current research is hyper-focused on performance metrics, with accuracy being the most dominant and consistently reported domain (143/247, 57.9%; mean 0.73). Although demonstrating technical capability is a necessary first step, this narrow focus comes at the expense of the foundational principles required for trustworthy AI. Critically, core domains such as safety, privacy, and security were almost entirely absent from the literature reviewed. This evaluation gap signifies a notable gap between the rapid growth in publications and the comprehensive, multidomain assessment needed before these technologies can be responsibly integrated into clinical practice. Importantly, the near-absence of privacy and security reflects a lack of evaluation and reporting rather than demonstrated poor performance; this review characterizes reporting patterns and does not audit deployed systems or clinical outcomes.

The issue of fairness and inclusiveness presents a noteworthy tension. Although it was the second most frequently assessed nonaccuracy domain (93/247, 37.7%), it consistently yielded the lowest average performance score (mean 0.58). This suggests that although the research community is beginning to acknowledge the importance of mitigating bias, current models still struggle considerably to generate equitable and inclusive outputs. Our analysis further revealed that this problem was especially pronounced in the surgical domain. Studies evaluating LLMs for surgical diagnoses reported a substantially lower mean fairness score (0.52) than those in internal medicine (0.61). This disparity may reflect underlying biases in surgical training data, demographic imbalances in procedural literature, or the nuanced language required for surgical decision-making, highlighting a critical area for targeted improvement.

The limited reporting and wide CIs for other essential categories, such as robustness, explainability, and verifiability, further underscore the immaturity of current evaluation methodologies. The high variability in these results indicates a lack of standardized metrics and a superficial approach to assessing model reliability.

Limitations

This study has several limitations that warrant consideration. Primarily, the LLM-assisted categorization was validated by a fully automated internal consistency check rather than against an independent human gold standard, so residual misclassification, particularly for low-prevalence domains, cannot be fully excluded. In place of a subjective risk-of-bias assessment, we used an objective RTI, which captures completeness of reporting rather than internal validity per se. This review was not prospectively registered (it was retrospectively registered on OSF), and screening was not independent or duplicated. Furthermore, our handling of metrics introduces other constraints. To respect authorial intent, we did not average multiple metrics from a single study, which may underrepresent the full scope of some evaluations. Additionally, normalizing diverse indicators, such as accuracy scores and Likert scales, onto a unified 0 to 1 scale oversimplifies their intrinsic meanings. The review is also susceptible to search and publication bias because it was limited to English-language literature and did not include unpublished research. Finally, given the rapid advancement of the LLM field, our findings should be interpreted as a snapshot in time, which may not reflect the latest technological or methodological developments.

Conclusions

In conclusion, the current evidence base is insufficiently comprehensive to support claims of deployment readiness for LLMs in high-stakes clinical environments. The rapid growth in accuracy-centered publications can create a misleading perception of technological readiness. To move forward, the research community must pivot from simplistic performance benchmarks to comprehensive, multidimensional evaluations grounded in established frameworks for trustworthy AI. Future studies must prioritize the development and application of robust metrics for safety, fairness, privacy, and reliability to help ensure that medical LLMs fulfill their promise without compromising patient well-being.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correspondence between the AI Guidelines for Business and health-specific AI frameworks (World Health Organization, Consolidated Standards of Reporting Trials–AI, Standard Protocol Items: Recommendations for Interventional Trials–AI, Developmental and Exploratory Clinical Investigation of Decision Support Systems Driven by AI, Fairness, Universality, Traceability, Usability, Robustness, and Explainability for AI, Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis+AI).
[DOCX File, 8 KB - ojphi_v18i1e97240_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Complete database-specific search strategies for the PubMed, Scopus, Web of Science, arXiv, and IEEE Xplore databases, including the search date (January 15, 2025) and records retrieved per database.
[DOCX File, 27 KB - ojphi_v18i1e97240_app2.docx]

Multimedia Appendix 3
Metric-family performance summaries and model-selection sensitivity analyses (best, mean, median, and primary models), including cluster bootstrap CIs and the preprint-exclusion sensitivity analysis.
[XLSX File, 12 KB - ojphi_v18i1e97240_app3.xlsx]

Multimedia Appendix 4
Prompt used for structured large language model–assisted data extraction, including the model (GPT-5 mini), access date (September 2025), and default decoding temperature.
[DOCX File, 10 KB - ojphi_v18i1e97240_app4.docx]

Multimedia Appendix 5
Category-keyword correspondence and coding manual for mapping extracted terms to trustworthy AI domains, including coding rules and borderline-case examples.
[DOCX File, 10 KB - ojphi_v18i1e97240_app5.docx]

Multimedia Appendix 6
Objective Reporting Transparency Index used as a fully automated surrogate for risk-of-bias assessment, including per-study index and higher-transparency sensitivity analysis.
[XLSX File, 15 KB - ojphi_v18i1e97240_app6.xlsx]

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 14 KB - ojphi_v18i1e97240_app7.docx]

References

1. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023 Mar 30;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
2. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature 2023 Aug;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
3. Mesko B. The role of artificial intelligence in precision medicine. Expert Rev Precis Med Drug Dev 2017;2(5):239-241. [doi: 10.1080/23808993.2017.1380516]
4. Wang B, Lai J, Cao H, et al. Enhancing the interoperability and transparency of real-world data extraction in clinical research: evaluating the feasibility and impact of a ChatGLM implementation in Chinese hospital settings. Eur Heart J Digit Health 2024 Sep;5(6):712-724. [doi: 10.1093/ehjdh/ztae066] [Medline: 39563908]
5. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
6. Shoja MM, Van de Ridder JM, Rajput V. The emerging role of generative artificial intelligence in medical education, research, and practice. Cureus 2023 Jun;15(6):e40883. [doi: 10.7759/cureus.40883] [Medline: 37492829]
7. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
8. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98. [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
9. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 1;27(3):491-497. [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
10. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health 2021 Nov;3(11):e745-e750. [doi: 10.1016/S2589-7500(21)00208-9] [Medline: 34711379]
11. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
12. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]

XSL•FO
RenderX

13. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
14. Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inform Decis Mak 2023 Jan 13;23(1):7. [doi: 10.1186/s12911-023-02103-9] [Medline: 36639799]
15. Waldock WJ, Zhang J, Guni A, Nabeel A, Darzi A, Ashrafian H. The accuracy and capability of artificial intelligence solutions in health care examinations and certificates: systematic review and meta-analysis. J Med Internet Res 2024 Nov 5;26:e56532. [doi: 10.2196/56532] [Medline: 39499913]
16. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. BMC Med Inform Decis Mak 2025 Mar 7;25(1):117. [doi: 10.1186/s12911-025-02954-4] [Medline: 40055694]
17. Asgari E, Montaña-Brown N, Dubois M, et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. NPJ Digit Med 2025 May 13;8(1):274. [doi: 10.1038/s41746-025-01670-7] [Medline: 40360677]
18. Moëll B, Sand Aronsson F. Harm reduction strategies for thoughtful use of large language models in the medical domain: perspectives for patients and clinicians. J Med Internet Res 2025 Jul 25;27:e75849. [doi: 10.2196/75849] [Medline: 40712151]
19. Campellone TR, Flom M, Montgomery RM, et al. Safety and user experience of a generative artificial intelligence digital mental health intervention: exploratory randomized controlled trial. J Med Internet Res 2025 May 23;27:e67365. [doi: 10.2196/67365] [Medline: 40408143]
20. Mahmoudi H, Chang D, Lee H, Ghaffarzadegan N, Jalali MS. Critical assessment of large language models' (ChatGPT) performance in data extraction for systematic reviews: exploratory study. JMIR AI 2025 Sep 11;4:e68097. [doi: 10.2196/68097] [Medline: 40934529]
21. Chelli M, Descamps J, Lavoué V, et al. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: comparative analysis. J Med Internet Res 2024 May 22;26:e53164. [doi: 10.2196/53164] [Medline: 38776130]
22. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020 Sep;26(9):1364-1374. [doi: 10.1038/s41591-020-1034-x] [Medline: 32908283]
23. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health 2020 Oct;2(10):e549-e560. [doi: 10.1016/S2589-7500(20)30219-3] [Medline: 33328049]
24. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ 2025 Feb 5;388:e081554. [doi: 10.1136/bmj-2024-081554] [Medline: 39909534]
25. Corrêa NK, Galvão C, Santos JW, et al. Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. Patterns (N Y) 2023 Oct 13;4(10):100857. [doi: 10.1016/j.patter.2023.100857] [Medline: 37876898]
26. AI guidelines for business ver 1.0 compiled. Ministry of Internal Affairs and Communications Japan; Ministry of Economy, Trade and Industry. 2024. URL: https://www.meti.go.jp/english/press/2024/0419_002.html [accessed 2026-08-10]
27. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. EU AI Act. URL: https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AIA-Final-Draft-21-January-2024.pdf [accessed 2026-08-06]
28. Artificial intelligence in software as a medical device. U.S. Food & Drug Administration. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device [accessed 2026-08-06]
29. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2026-08-06]

Abbreviations

AUC: area under the curve
CLAIM: Checklist for Artificial Intelligence in Medical Imaging
CONSORT: Consolidated Standards of Reporting Trials
FDA: Food and Drug Administration
FUTURE: Fairness, Universality, Traceability, Usability, Robustness, and Explainability
LLM: large language model
ML: machine learning
OSF: Open Science Framework
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
RTI: Reporting Transparency Index
SMOG: Simple Measure of Gobbledygook
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
STARD: Standards for Reporting Diagnostic Accuracy Studies

XSL•FO
RenderX

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
WHO: World Health Organization

Edited by E Mensah; submitted 05.Apr.2026; peer-reviewed by J Mino-Ayala, J Vrdoljak, X Tao; revised version received 22.Jun.2026; accepted 27.Jul.2026; published 20.Aug.2026.

Please cite as:
Matsuoka H, Takahashi T, Semitsu T, Matsuoka K, Hayashi R
Safety-Oriented Evaluation of Large Language Models in Health Care: Guideline-Informed Systematic Review
Online J Public Health Inform 2026;18:e97240
URL: https://ojphi.jmir.org/2026/1/e97240
doi:10.2196/97240

© Hikaru Matsuoka, Takayuki Takahashi, Takayuki Semitsu, Keiko Matsuoka, Ryoma Hayashi. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 20.Aug.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating Methodological Coherence and Evidence Recognition in Digital Health Systematic Reviews: Sample-based Meta-research Study

Uwe Buddrus[1], BA; Jan-David Liebe[1,2], Dr rer pol

[1]Professorship Digital Society, Faculty of Business, Economics and Social Sciences, Hochschule Osnabrück, CO-Gebäude, Raum 102, Albrechtstraße 30 a, Osnabrück, Germany

[2]Institute of Medical Informatics, UMIT - Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik, Hall, Tyrol, Austria

Corresponding Author:
Jan-David Liebe, Dr rer pol
Professorship Digital Society, Faculty of Business, Economics and Social Sciences, Hochschule Osnabrück, CO-Gebäude, Raum 102, Albrechtstraße 30 a, Osnabrück, Germany

Abstract

Background: Despite a growing number of systematic reviews on digital health interventions, many do not sufficiently support the recognition of conclusive evidence. Methodological shortcomings may impede the identification and communication of robust findings. Abstracts are the basis for study selection in systematic reviews and are increasingly used in automated screening processes and rapid assessments.

Objective: This meta-research study examines to what extent systematic reviews apply methodological standards—particularly the specification of PICO (population or problem, intervention, comparison, and outcome) elements—and how this relates to the likelihood of conclusive evidence recognition. It is based on a random sample and focuses on the assessment at the abstract level, as abstracts are used independently of the specific review choice to screen and select studies for evidence synthesis, making them critical for evidence recognition.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we conducted a comprehensive database search (2011 - 2023). From 2528 eligible systematic reviews, a random sample of 250 abstracts was analyzed descriptively. Abstracts were assessed for PICO specification and evidence conclusiveness in the context of further study characteristics.

Results: In total, 48% (119/250) of reviews showed low or very low PICO specification, and 64% (159/250) reported inconclusive or weak evidence. Higher specification of outcomes and problems was moderately associated with conclusive evidence. Beside the formulation of the research question along the PICO scheme, we identified recurring issues in search and screening strategy design (eg, limited database use, vague search terms, and long search periods), restrictive eligibility criteria (eg, exclusive reliance on randomized controlled trials), inconsistent use of quality appraisal tools, and underusage of alternative synthesis methods to hinder evidence recognition.

Conclusions: Our findings suggest that methodological coherence across all review stages is a necessary condition to ensure conclusions and evidence-informed decisions in the digitalization of health care are both valid and meaningful. A structured PICO-based framework, which is aligned with current research, builds on well-established categories and provides clear and differentiated definitions that may enhance the focus and evidentiary strength of future reviews.

(Online J Public Health Inform 2026;18:e78210) doi:10.2196/78210

KEYWORDS

systematic reviews; digital health interventions; evidence; outcomes; effects; PICO; title and abstract screening; evidence recognition; methodological flaws; population or problem, intervention, comparison, and outcome

Introduction

The digital transformation of health care has led to a substantial increase in studies evaluating the effects of digital health interventions (DHIs) [1,2]. Despite this growing body of research, decision-makers still lack conclusive and consistent evidence on the effects of DHIs. Systematic reviews, although considered the gold standard for evidence synthesis, frequently fail to provide robust conclusions. This discrepancy raises fundamental questions about the methodological quality and scientific rigor of such reviews. Given the role of systematic reviews in shaping digital health policy and clinical guidance, it is crucial to understand how methodological shortcomings affect the reliability of their conclusions.

XSL•FO
RenderX

Poorly formulated research questions and ineffective evidence search and screening strategies were previously identified to impede the retrieval and synthesis of relevant evidence [3].

The PICO (population, problem, or patient, intervention, comparison, and outcome) framework has been widely adopted in clinical research to formulate answerable questions and design efficient search strategies [3-6]. Empirical studies confirm that applying PICO is predominant in clinical practice for structuring questions that compare one intervention with an alternative [7], enhances the precision of literature searches, and supports the identification of relevant studies [8]. Using PICO is recommended for interventional effectiveness and economic cost-effectiveness studies [6], especially pertinent for analyzing DHI effects.

It is evident that PICO is critical in the research process, including the formulation of the research questions, keyword development for the search strategy, and the definition of eligibility criteria. Nonetheless, its application in systematic reviews appears to be inconsistent and often not rigorous in these early stages of the research process. This may limit the capacity of systematic reviews to deliver actionable conclusions, especially in a field as heterogeneous and dynamic as digital health. Thus, we consider PICO as a methodological key factor and hypothesize that low PICO specification correlates with low evidence recognition.

To date, no comprehensive analysis has quantified the extent and specificity of the use of PICO elements in systematic reviews of DHIs. Furthermore, the relationship between the degree of PICO specification and the likelihood of finding conclusive evidence remains underexplored.

To address these gaps, this study investigates the extent to which systematic reviews on DHIs specify PICO elements and examines whether this level of specification is associated with the likelihood of reporting conclusive evidence.

Given that abstracts are the basis for study selection in systematic reviews and are increasingly used in automated screening processes and rapid assessments, abstracts serve as a primary information source for evidence identification. This study, therefore, focuses on the abstract level to evaluate how methodological structure is communicated and how it may influence the perceived strength of evidence when screening and selecting studies for full-text review.

The following research questions guided our analysis:

1. How frequently and to what degree are the PICO elements specified in the abstracts of systematic reviews and meta-analyses on DHI effects?
2. What is the relationship between the specification of PICO elements and the availability of conclusive evidence?
3. How are methodological study characteristics (eg, search period, number of databases searched, included study types, and number of included studies) associated with the likelihood of conclusive evidence in systematic reviews on DHIs?

Methods

Overview

We conducted this meta-research study following selected principles of the scoping review methodology according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [9], for example, broad inclusion criteria and descriptive synthesis. We performed a structured, quantitative content analysis on a random sample of 250 abstracts from systematic reviews and meta-analyses researching the effects of DHIs.

Table 1 provides an overview of the key aspects in each methodological step.

XSL•FO
RenderX

Table . Summary overview of methodological steps.

Methodological step	Description
Search strategy	• Systematic reviews or meta-analyses • A broad range of DHIs[a] • A broad range of outcomes • In any health care setting • From 2011 to October 2023 • In 5 leading databases
Study selection	• Title and abstract screening in reviewer pairs • 10 inclusion and 6 exclusion criteria • Random sample of 250 abstracts for 95% CI and 6% margin of error
Data extraction	• Exclusively from titles and abstracts • 16 data extraction fields and variables
Data analysis	• Transformation into metric, nominal, and ordinal variables • Univariate quantitative descriptive analysis of study characteristics, and representation of PICO[b] elements and criteria for conclusive evidence • Bivariate and correlation analysis using Spearman rank correlation coefficient (ρ) of PICO specification and evidence

[a]DHI: digital health intervention.

[b]PICO: population or problem, intervention, comparison, and outcome.

Search Strategy

The search strategy was developed with the aim to include existing systematic reviews and meta-analyses which address (1) a broad range of DHI technologies and (2) a broad range of potential associated outcomes in (3) any health care setting.

A total of 33 DHI-related and 41 outcome-related search terms were used and—depending on the database searched—applied as Medical Subject Headings terms to subject headings or as plain search terms on titles and abstracts. A complete documentation of all search terms is available in Multimedia Appendix 1.

A comprehensive literature search was conducted for the time frame 2011 to October 2021 for a preceding umbrella review project, which was updated in October 2023 for more in-depth research.

To capture a large heterogeneity of care settings and DHI application scenarios, the search included all areas of health care—from inpatient hospital care through outpatient and community-based care to self-care, without any geographic boundaries.

The search was limited to systematic reviews or meta-analyses published in English or German.

We searched 5 leading databases—Scopus, AISeL, EBSCO or CINAHL, Cochrane Library, and MEDLINE via PubMed.

Study Selection

The search results were uploaded to Covidence (Veritas Health Innovation Ltd), a software for managing and streamlining systematic reviews. All potential studies for further analyses, such as evidence and gap mapping, were screened in pairs by UB and Jan-Oliver Kutza, and Johannes Thye and Moritz Esdar, and discrepancies were resolved through discussion or consultation with a senior reviewer (JDL or Ursula Hübner). A complete documentation of all inclusion and exclusion criteria is available in Multimedia Appendix 2.

The initial abstract review and screening process was completed in February 2024. After all abstracts were screened, 2528 eligible studies for subsequent analyses were exported to Microsoft Excel, additional duplicates were identified, checked, and omitted, and a random sample of 250 studies was drawn, using the random number assignment function in Excel, by the lead author (UB). The sample size of 250 abstracts was chosen as the basis for the detailed data extraction and quantitative analysis, to represent about 10% (250/2528) of eligible studies, and to have a 95% confidence level that the actual value is within ±6% of the measured or surveyed value when exploring potential relationships between study characteristics reported in the abstracts and the likelihood to find conclusive evidence in the full texts.

Data Extraction

In line with our research questions, a set of 16 data extraction fields and variables was developed. Next to meta-information on study design (particularly if meta-analysis was performed and randomized controlled trial [RCT]–focus), publication year, search period in years, databases searched, the total number of studies identified by the search strategy, the number of included studies, the journal, and statements regarding evidence quality and quality assessment tools, these were the descriptions related to the PICO elements problem and population, intervention, comparator, and outcome, as well as the results and conclusions, representing the scope of evidence.

All information was exclusively identified and extracted from titles and abstracts.

XSL•FO
RenderX

This decision was based on the relevance of abstracts as the main communicative interface in the screening process for evidence synthesis and decision-making contexts. Abstracts are increasingly used in automated screening and mapping tools, making them a critical unit of analysis.

Data Analysis

Data Transformation and Classification

The quantitative analysis focused on the meta-information of study characteristics and the representation of PICO elements and criteria for conclusive evidence in abstracts of systematic reviews and meta-analyses. To facilitate quantification, extracted data were transformed into metric (eg, search period in year, number of databases searched, and cumulative PICO score), nominal (eg, yes or no for meta-analysis and RCT-focus), and ordinal variables (eg, search period: short=1‑10 y, medium=11‑20 y, and long=21+ y; number of databases: high=7+, medium=4‑6, and low<4; and number of studies included: high=21+, medium=11‑20, and low=1‑10).

The ordinal classification scheme for PICO elements and likelihood of conclusive evidence was developed through multiple iterations, based on a qualitative review of the wording used in titles, objectives, methods, results, and conclusions of the abstracts. Additionally, the classification of the likelihood of conclusive evidence was informed by literature. The criteria were refined until they could be unambiguously applied to every abstract in the sample. The resulting ordinal variables for the classification of PICO elements were high=3, medium=2, low=1, and not specified=0, and for the likelihood of conclusive evidence high=3, medium=2, low=1, and inconclusive=0. All criteria accompanied by examples from extracted data are available in the Multimedia Appendix 3. To ensure clarity, consistency, and applicability, the scheme was pilot tested on 20 abstracts by UB and JDL.

The overall level of PICO specification was calculated as a cumulative PICO score (minimum 0 and maximum 12), whereby a high level of specification for any PICO element was attributed a score of 3, medium=2, low=1, and not specified=0. Since 2 P-elements (P1 problem and P2 population) were considered, the average of both scores was calculated to avoid overweighting of the P-element.

Quantitative Descriptive and Correlation Analysis

We performed univariate analysis on the variables for study characteristics, the classification of PICO elements, and the classification of the likelihood for conclusive evidence, creating descriptive statistics, including frequency tables. Where available, metric variables (eg, cumulative PICO scores and publication year) were used in the analysis and complemented by group comparisons of their ordinal counterparts (eg, for search periods, databases, and included studies). Nominal variables (eg, meta-analysis and RCT focus) were handled appropriately using group comparisons.

Associations between PICO specification and the likelihood of conclusive evidence were explored by bivariate analysis resulting in cross-tabulations and box plots. To test potential correlations statistically, we used Spearman rank correlation coefficient (ρ) due to the ordinal nature of key variables. *P* values were reported for completeness, but following Andrade [10], interpretation emphasized effect sizes and directional trends over strict significance thresholds. Correlation strength was categorized based on the conventions described by Kuckartz et al [11].

Results

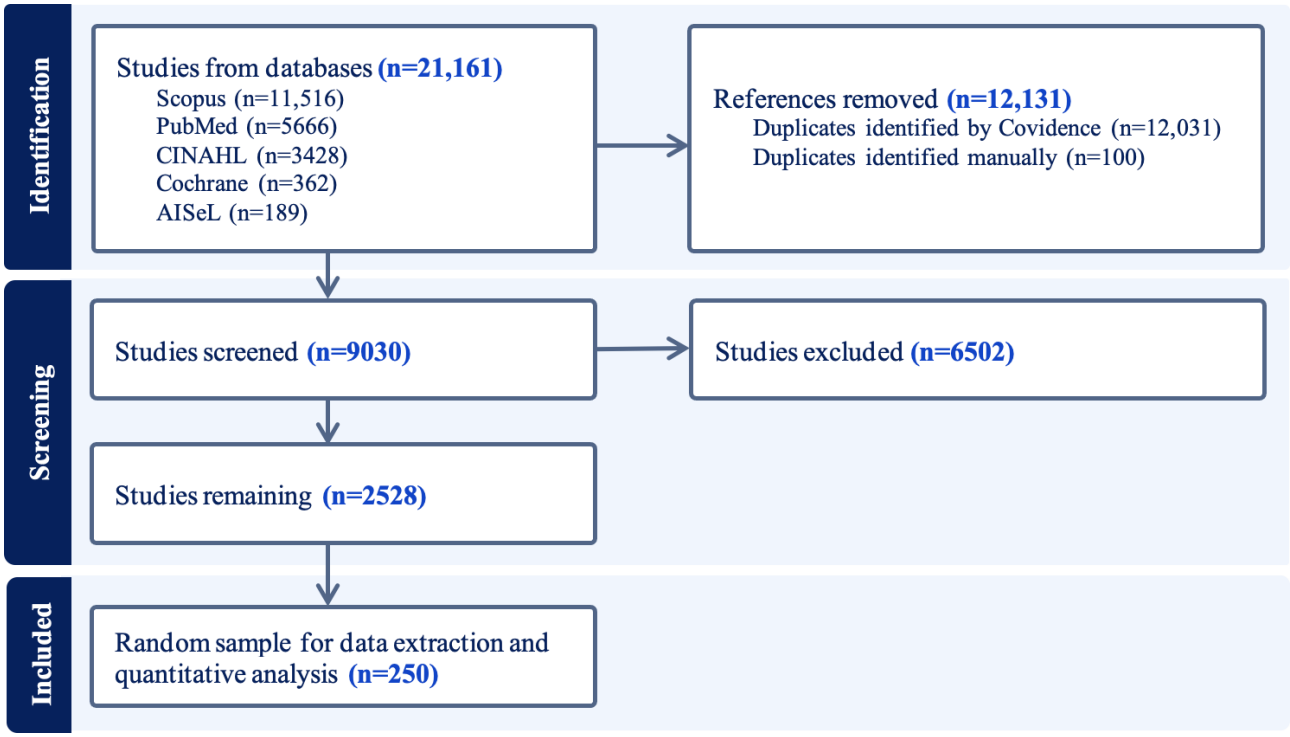
PRISMA Flowchart

Out of 21,161 initial records, 2528 studies remained after screening. From these, a random sample of 250 was drawn for detailed analysis.

Figure 1 shows the PRISMA flowchart.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Study Characteristics in the Random Sample

We analyzed 9 study characteristics for the discussion in the context of their influence on evidence recognition and indications of possible methodological deficits.

Table 2 provides an overview, showing the number (n) of the 250 abstracts in which a study characteristic was reported together with the relative shares (%).

XSL•FO
RenderX

Table . Summary overview of study characteristics in the sample.

Study characteristics	Studies, n (%)
Study design	
Reported, of which:	250 (100)
Systematic review	146 (58.4)
Systematic review and meta-analysis	58 (23.2)
Systematic literature review	23 (9.2)
Meta-analysis	13 (5.2)
Systematic review of (systematic) reviews	6 (2.4)
Systematic review of systematic reviews and meta-analyses	4 (1.6)
Study types included	
Reported, of which:	170 (100)
RCTs[a] only	62 (36.5)
RCTs among others	44 (25.9)
No RCTs	64 (37.6)
Not reported, n (% of 250)	80 (32)
Publication year	
Reported, of which:	250 (100)
2023	42 (16.8)
2022	59 (23.6)
2021	32 (12.8)
2020	21 (8.4)
2011 - 2019	96 (38.4)
Search period in years	
Reported, of which:	159 (100)
1 - 10	29 (18.2)
11 - 20	32 (20.1)
>20 (eg, from inception)	98 (61.6)
Not reported, n (% of 250)	91 (36.4)
Number of databases searched	
Reported, of which:	202 (100)
1 - 3	67 (33.2)
4 - 6	102 (50.5)
7 - 14	33 (16.3)
Not reported, n (% of 250)	48 (19.2)
Databases used (top 5)	
Reported, of which:	179 (100)
PubMed	110 (61.5)
Embase	107 (59.8)
MEDLINE (including Medline through PubMed)	102 (57)
CINAHL (EBSCO)	62 (34.6)
Web of Science	58 (32.4)
Not reported, n (% of 250)	80 (32)
Number of studies retrieved and screened	

XSL•FO
RenderX

Study characteristics	Studies, n (%)
Reported, of which:	82 (100)
1 - 1000	28 (34.1)
1001 - 2000	21 (25.6)
2001 - 3000	10 (12.2)
3001 - 4000	8 (9.8)
>4000	15 (18.3)
Not reported, n (% of 250)	168 (67.2)
Number of studies included	
Reported, of which:	235 (100)
1 - 10	65 (27.7)
11 - 20	75 (31.9)
21 - 30	42 (17.9)
31 - 40	23 (9.8)
>40	30 (12.8)
Not reported, n (% of 250)	15 (6)
Critical appraisal and RoB[b] tools (top 5)	
Reported, of which:	76 (100)
Cochrane Collaboration Risk of Bias tools	27 (32.5)
GRADE[c]	10 (12)
Risk of bias assessment tools (unspecified)	7 (8.4)
Joanna Briggs Institute Critical Appraisal Checklist	6 (7.2)
AMSTAR[d] or AMSTAR-2	5 (6)
Not reported, n (% of 250)	174 (69.6)

[a]RCT: randomized controlled trial.

[b]RoB: risk of bias.

[c]GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

[d]AMSTAR: A Measurement Tool to Assess Systematic Reviews.

- Study design: 30% (75/250) feature meta-analyses, 4% (10/250) were systematic reviews of (systematic) reviews.
- Study types: Overall, 32 study designs were mentioned in 170 abstracts; RCTs in 106 (62.4%), followed by controlled clinical trials (in 13, 7.6%), and pre-post studies (in 13, 7.6%), nonrandomized controlled trials in 12 (7.1%), and cohort studies in 10 (5.9%).
- Publication year: Over the last years, publications increased exponentially; 53.2% of studies (133/250) were published since 2021.
- Search period in years: Search periods were generally rather long, and “from inception” of the databases was often (27/159, 17%) specified.
- Number of databases searched: Just over half of the studies were based on the search in 4 to 6 databases, with a mean of 5 (SD 2) and a median of 4 (IQR 3-6). One-third of searches were performed in just 1-3 databases.
- Databases used: Overall, 65 databases were mentioned in 179 abstracts; 9 were mentioned in more than 10% of abstracts, 27 only once (1/179, 0.6%).
- Number of studies retrieved and screened: Researchers retrieved and screened between 49 and 42,946 studies with a median of 1491 (IQR 588 - 3389) and a mean of 3016 (SD 5502).
- Number of studies included: Despite generally large volumes of studies screened, relatively small shares were included (median 1.2%, IQR 0.6 - 3.1; mean 4.2%, SD 8%), ranging overall between 0.02% and 44.9%. The absolute numbers range from 1 to 236, with a median of 17 (IQR 10-30) and a mean of 25 (SD 29).
- Critical appraisal and risk of bias (RoB) tools: Overall, 29 tools were mentioned in 76 abstracts; 2 tools in 18 abstracts (23.7%). In this context, it is noteworthy that, if reported in the abstract, study or evidence quality was often described as low (19/42, 45%) and RoB as high (11/23, 48%).

Level of Specification of PICO Elements

The level of specification of PICO elements was analyzed in the context of the quality of current research questions in the field of DHIs. Each PICO element plays a distinct role in

XSL•FO
RenderX

framing the scope and analytical precision of a review. Their individual specifications were therefore examined separately.

Table 3 and Figure 2 display the level of specification of PICO elements according to the classification scheme (Table S1 in Multimedia Appendix 3) across the sample. A key finding is that no single PICO element was highly specified in more than half of the abstracts.

XSL•FO
RenderX

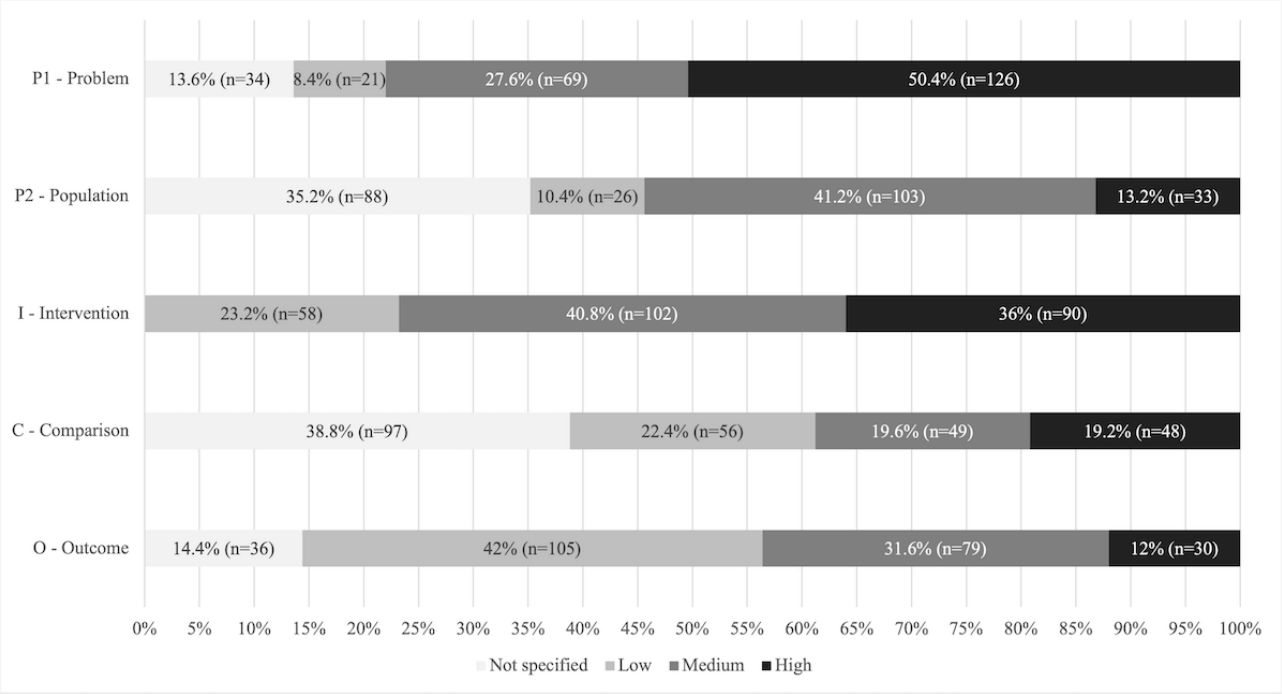
Table . Level of specification of PICO[a] elements, criteria for classification.

PICO element, classification	Criteria for classification	Studies (N=250), n (%)
P1 (Problem: disorder or condition)		250 (100)
High	One medical problem or objective is clearly specified	126 (50.4)
Medium	A relatively homogenous group of medical problems or objectives	69 (27.6)
Low	Multiple highly heterogenous medical problems or objectives	21 (8.4)
Not specified	No specification whatsoever	34 (13.6)
P2 (Population: user group)		250 (100)
High	One, clearly specified, eg, by disorder AND additional characteristics	33 (13.2)
Medium	A relatively homogenous population or target group, eg, patients	103 (41.2)
Low	Multiple highly heterogenous populations or target groups	26 (10.4)
Not specified	No specification whatsoever	88 (35.2)
I (Intervention: technology)		250 (100)
High	One specific technology or aspect of homogenous technologies	90 (36)
Medium	A relatively homogenous group of technologies	102 (40.8)
Low	An unspecific range of highly heterogenous technologies	58 (23.2)
Not specified	No specification whatsoever	0 (0)
C (Comparison: comparable clinical practice or setting)		250 (100)
High	One care setting is clearly specified	48 (19.2)
Medium	A relatively homogenous group of care settings or explicit focus on intersectoral exchange	49 (19.6)
Low	An unspecific range of highly heterogenous care practices or settings, which may include no care, self-care, and multiple healthcare settings	56 (22.4)
Not specified	No specification whatsoever	97 (38.8)
O (Outcome)		250 (100)
High	One outcome is clearly specified with defined indicators	30 (12)
Medium	A relatively homogenous group of outcomes or a small number of specified and predefined but diverse outcomes of interest	79 (31.6)
Low	An unspecific range of heterogenous outcomes	105 (42)
Not specified	No focus on outcomes; outcomes are one of multiple objectives	36 (14.4)

[a]PICO: population, problem, or patient, intervention, comparison, and outcome.

XSL•FO
RenderX

Figure 2. Level of specification of PICO (population, problem, or patient, intervention, comparison, and outcome) elements (% and n of sample, N=250).



Research focus on 1 medical problem (P1), such as the disease, condition, or clinical objective, was highly specified in 50.4% (126/250) of the abstracts. The intervention (I), referring to the technology used, was specified to some degree in all abstracts, with levels of specification distributed from low (58/250, 23.2%) to high (90/250, 36%). Outcomes (O) were relatively well-defined in 43.6% (109/250) of abstracts, whereas the remaining lacked a clear focus on predefined outcomes. Notably, 35.2% (88/250) of abstracts did not specify the population (P2), referring to the user group of the intervention, and 38.8% (97/250) lacked any specification for the comparison (C) element, referring to the health care setting applicable to the intervention. The population and comparison were highly specified in only 13.2% (33/250) and 19.2% (48/250) of abstracts, respectively.

Level of PICO Specification and Likelihood of Conclusive Evidence

We hypothesized that low PICO specification in research objectives, methods, and questions is associated with low evidence recognition. Therefore, our study aimed to examine the relationship between PICO specification and the availability of conclusive evidence described in the abstracts.

To this end, criteria for the likelihood of conclusive evidence were defined based on literature and a qualitative review and thematic analysis of all results and conclusions reported in the abstracts (Table S2 in Multimedia Appendix 3).

Table 4 provides a summary of these criteria and the results of their application to the sample.

Table . Likelihood of conclusive evidence, criteria for the classification.

Likelihood of conclusive evidence, classification	Criteria for classification	Studies (N=250), n (%)
High	Meta-Analysis AND multiple (>1) “significant” results reported	19 (7.6)
Medium	Meta-Analysis AND at least comparable results to usual care OR no meta-analysis, but at least one “significant” effect reported for >25% of relevant studies	72 (28.8)
Low	Neither meta-analysis nor any “significant” results reported. Evidence is mainly or only reported qualitatively. If any quantitative results are reported, none is described as “significant.”	135 (54)
Inconclusive	Evidence is not reported at all OR without any directional conclusion OR although studies report positive evidence, also quantitatively, there are equally contradictive effects reported.	24 (9.6)

To assess whether methodological focus correlates with result strength (our second research question), we examined the relationship between the overall level of PICO specification, based on a cumulative PICO score (minimum=0, maximum=12;

XSL•FO
RenderX

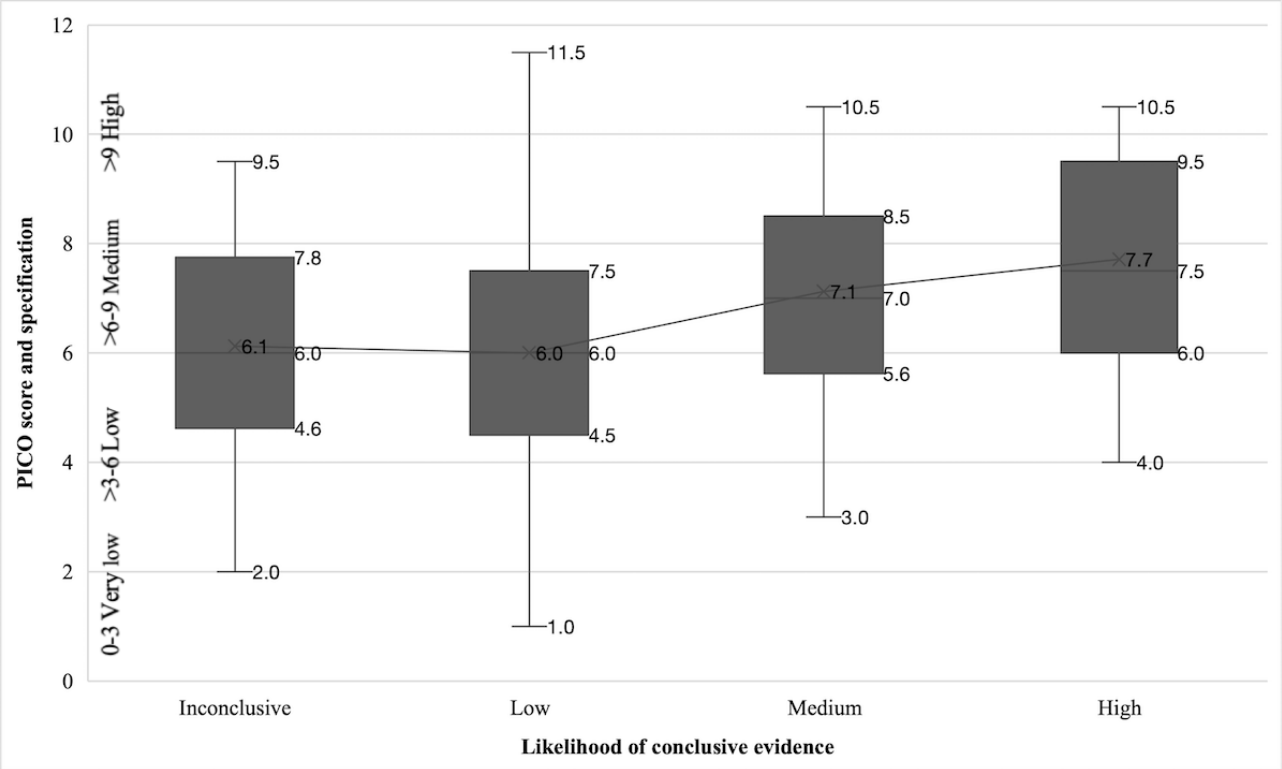
0 - 3=very low, >3 - 6=low, >6 - 9=medium, and >9=high), and the likelihood of conclusive evidence being reported. Table 5 and Figures 3 and 4 show the results.

Table . Correlation between overall PICO[a] specification and likelihood of conclusive evidence (n [%] of sample, N=250).

Overall PICO specification	Likelihood of conclusive evidence				
	High	Medium	Low	Inconclusive	Total
High	6 (31.6)	11 (15.3)	10 (7.4)	3 (12.5)	30 (12)
Medium	8 (42.1)	38 (52.8)	48 (35.6)	7 (29.2)	101 (40.4)
Low	5 (26.3)	22 (30.6)	64 (47.4)	12 (50)	103 (41.2)
Very low	0 (0)	1 (1.4)	13 (9.6)	2 (8.3)	16 (6.4)
Total	19 (100)	72 (100)	135 (100)	24 (100)	250 (100)

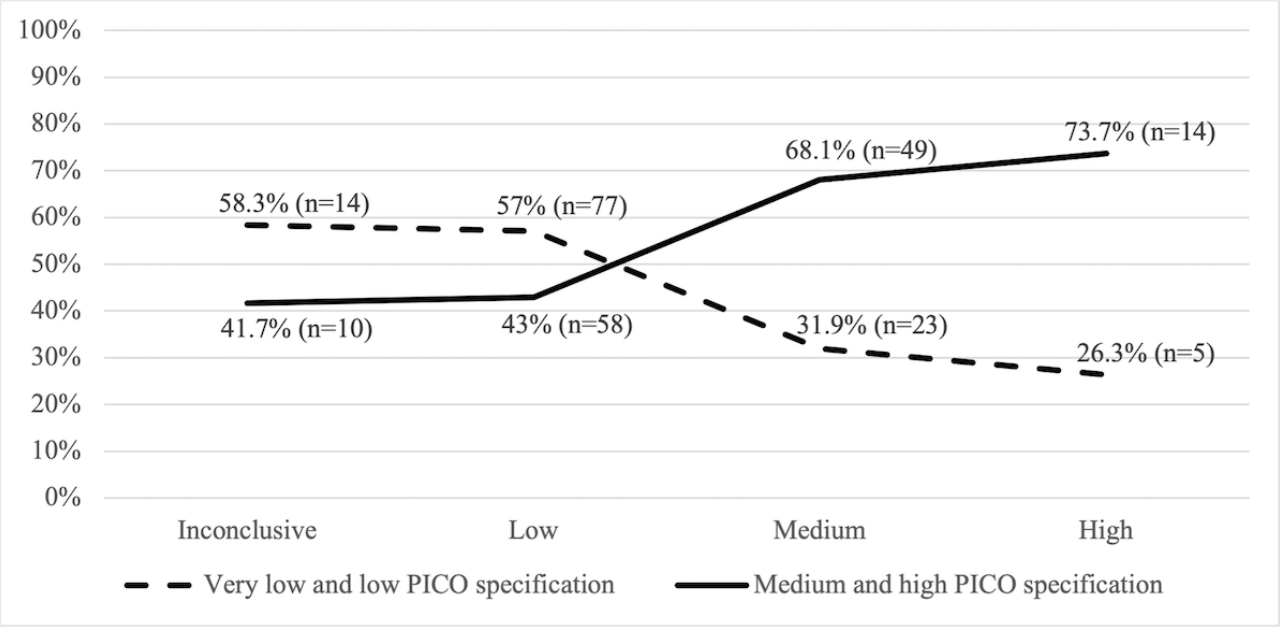
[a]PICO: population, problem, or patient, intervention, comparison, and outcome.

Figure 3. Distribution of the cumulative PICO score by likelihood of conclusive evidence classification. PICO: population, problem, or patient, intervention, comparison, and outcome.



XSL•FO
RenderX

Figure 4. Combined PICO specification shares by likelihood of conclusive evidence classification, % (n) of abstracts in each evidence class. PICO: population, problem, or patient, intervention, comparison, and outcome.



Among the 30 abstracts (30/250, 12%) with a high PICO score, only 4 specified all 4, 20 specified 3, and 6 specified only 2 PICO elements highly. Of those 101 with a medium PICO score, only 6 specified 3, and 46 specified 2 PICO elements highly. The 103 abstracts with a low PICO score included only 8 with 2 PICO elements and 44 with 1 PICO element being highly specified.

This means that very few systematic reviews (4/250; 1.6%) have a clear focus on 1 problem or population, 1 intervention, 1 comparative setting, and 1 outcome. On the other hand, 47.6% (119/250) demonstrate a low or very low level of PICO specification, indicating lack of research focus and a very broad scope.

Only 6 abstracts in the sample reported highly conclusive positive evidence on highly specified PICO-based research questions (6/19, 31.6% of abstracts indicating highly conclusive evidence; 6/250, 2.4% of the entire sample).

Figure 3 and Figure 4 visualize the increasing likelihood of conclusive evidence with rising PICO specification. In Figure 3, PICO specification is represented by the PICO score.

Figure 4 illustrates the distribution of the combined shares for very low and low, and medium and high PICO specification within each evidence class, highlighting that methodological clarity alone does not guarantee robust findings.

While the combined medium and high PICO specification is associated with more robust findings, about two-fifths of the abstracts that remained inconclusive or yielded a low likeliness of conclusive evidence did so regardless of the medium or high level of PICO specification.

Correlation analysis confirmed a weak to medium positive relationship with statistical significance ($\rho=0.26$; $P<.001$) between overall PICO specification (ie, PICO score) and conclusive evidence. Higher specification of the outcome of interest (PICO element O; $\rho=0.27$; $P<.001$) and of the problem (PICO element P1; $\rho=0.17$; $P=.008$) were most associated with conclusive evidence being reported. In contrast, specification of the PICO elements population (P2), intervention (I), and comparison (C) showed no or only weak relationships with conclusive evidence.

Methodological Characteristics and Likelihood of Conclusive Evidence

To illustrate the implications for future review practice, we have aligned the key insights of this study with the typical phases of a systematic review process, adapted from the core methods of conducting a Cochrane review [12,13].

Figure 5 highlights where critical flaws most frequently occur and also offers targeted suggestions to strengthen methodological coherence and improve evidence recognition, particularly in the context of complex interventions, such as DHIs.

XSL•FO
RenderX

Figure 5. Overview of key findings and potential flaws, and recommendations along the steps of a systematic review. DHI: digital health intervention; EGM: evidence and gap mapping; PICO: population, problem, or patient, intervention, comparison, and outcome; RCT: randomized controlled trial.

XSL•FO
RenderX

In the interpretation of the above data, it should be noted that sample sizes (n) on which percentages are based vary. Refer to Tables 2, 4, and 5 and Multimedia Appendix 4 for details.

Research question: Our findings indicate inadequate application of PICO in formulating research questions, with 47.6% (119/250) of abstracts displaying low or very low PICO specification, and 63.6% (159/250) featuring poor or inconclusive evidence.

Although statistically significant, the correlation between PICO specification and conclusive evidence remains modest (ρ=0.26; *P*<.001). Outcome and problem specification showed the strongest associations with conclusive evidence, indicating their central role in the methodological quality of systematic reviews. Despite this apparent relationship, more than 40% of abstracts with poor (58/135) or inconclusive (10/24) evidence exhibit medium to high PICO specification.

Eligibility criteria: Currently, 36.5% (62/170) of systematic reviews rely exclusively on RCTs despite the availability of several alternatives, such as pre-post studies, nonrandomized controlled trials, cohort studies, time series, and case studies, that may also be appropriate, suggesting a need to expand the scope of study designs considered.

Search strategy: Despite more than 65 databases being available, 33.2% (67/202) of authors only used 3 or fewer databases. The number of studies retrieved and screened varies greatly, indicating significant variation in search strategy quality, particularly search periods and the formulation of search terms, operators, and search strings. Interestingly, no apparent relationship exists between PICO specification levels in the abstracts and search results. In total, 6 of the 10 lowest search results with less than 150 identified studies show poor PICO specification, while 6 of the 10 highest search results are associated with a high level of PICO specification.

Screening criteria: Despite generally large numbers of studies being identified and screened, only a small percentage (median 1.2%, IQR 0.6% - 3.1%; mean 4.2%, SD 8%) and absolute number (median 17; mean 25) is finally included, with 27.7% (65/235) of systematic reviews including 10 or less studies. In this context, it is noteworthy that 24.8% (62/250) of systematic review abstracts are not structured along background, methods, results, and conclusions, with missing details having a potential impact on the screening decision.

Critical appraisal tools: There is a host of tools available for different study designs (eg, systematic reviews, mixed methods, and primary research), research focus (eg, care practice, diagnostic accuracy, and health economics), and appraisal focus (eg, RoB and quality), of which 29 were identified in the sample. The choice of appraisal tool and the extreme variability in their criteria can significantly influence the exclusion of relevant studies, especially due to harsh appraisal ratings. Even for included studies, it is noteworthy that, if reported in the abstract, study or evidence quality is often described as low (19/42, 45%) and RoB as high (11/23, 48%). On the other hand, Blum et al [14] found no association with study quality for their results.

Evidence synthesis: Meta-analysis, the most desirable output of a systematic review to derive conclusive evidence, is featured in only 30% (75/250) of studies. Moreover, 13.2% (33/250) of abstracts include explicit comments on considerable heterogeneity or variability of study designs, populations, interventions, and outcomes, often stating that this prevented meta-analysis [15-25]. Alternative synthesis methods, such as thematic analysis [26] or statistical result aggregation [27,28], are underused. Higher-level syntheses, such as umbrella reviews of systematic reviews, are still rare (10/250, 4%).

Conclusion validity: In total, 54% (135/250) of systematic review abstracts reported neither meta-analysis nor any significant results; 9.6% (24/250) reported no evidence at all or no directional conclusion or equally contradictive effects. Although our analysis did not delve deeply into outcome effects, we stress the importance of selecting outcomes sensitive to change, as noted by Shen et al [29].

Discussion

Summary of Main Findings

This meta-research study aimed to identify key methodological factors that limit conclusive evidence recognition in systematic reviews on DHIs. Our findings show that the overall level of PICO specification in the abstracts was low or very low in nearly half of the sample, and that two-thirds of the reviews yielded inconclusive or weak evidence. Although higher levels of PICO specification, particularly regarding outcomes and clinical problems, were moderately associated with the likelihood of conclusive findings, these factors alone were not sufficient to guarantee evidence clarity.

This observation supports our initial assumption that PICO-structured research questions provide an important methodological foundation. However, it also reveals the limitations of assuming a linear relationship between methodological structure and evidentiary strength. Indeed, about two-fifths of reviews with low or inconclusive evidence had medium to high PICO specification. This suggests that even clearly defined research questions may fail to translate into robust evidence if other aspects of the review process are flawed.

Among these aspects, we identified recurring issues in search strategy design (eg, limited database use and vague search terms), restrictive eligibility criteria (eg, exclusive reliance on RCTs), and inconsistent use of quality appraisal tools. These elements interact with question formulation and may explain why even well-specified PICO frameworks can fail to yield strong evidence. Notably, substantial heterogeneity in study designs and outcomes further impedes evidence synthesis, particularly when meta-analyses are not feasible or misapplied.

Detailed Discussion

Our findings underscore the need for a more integrated methodological approach. Rather than focusing solely on formal structures, review authors should align research questions, search strategies, eligibility and screening criteria, critical appraisal, evidence synthesis methods, and conclusions to the specific characteristics of digital health technologies. This is particularly relevant for DHIs, which often involve complex, context-sensitive mechanisms of action that challenge conventional evaluation models.

XSL•FO
RenderX

Research question: While high PICO specification does not guarantee conclusive results, a well-defined scientific starting point is critical [30]. In formulating the research question, both too specific and too broad specification of PICO elements appear more likely to lead to inconclusive results. We conclude, at a medium level of specification, that is, clearly defined and differentiated categories, in at least 2 or 3 PICO elements, would be optimal for many systematic reviews.

Eligibility criteria: Broad inclusion criteria, represented by low PICO specification, relate to substantial heterogeneity [30]. Conversely, narrow inclusion criteria, such as focusing solely on RCTs, also add challenges. Despite RCTs being the gold standard for intervention evaluations [31], they might not be fully appropriate in evaluating DHIs, due to long time frames, high costs, rigid protocols, and DHI specificities, such as individual tailoring and interaction effects that RCTs can insufficiently address [32-34].

Search strategy: A comprehensive and sensitive search strategy in multiple databases is recommended [30]. Our findings highlight the necessity for more methodological rigor in comprehensive search strategies to mitigate the inconsistent retrieval of relevant studies. This concerns particularly searches across multiple databases and search terms and strings that are more aligned with the research questions. Exponential growth in the number of publications and increasing quality of studies over time leads us to suggest restricting search periods for future systematic reviews to a maximum of 10 years.

Screening criteria: Our findings suggest that screening criteria might not effectively capture all relevant research to ensure comprehensive evidence synthesis, necessitating a reassessment in the screening process. Given the limitations observed with broad and narrow eligibility criteria, more consideration of inclusion and exclusion criteria, especially a focus on RCTs, seems to be required to balance comprehensiveness and specificity.

While the abstract structure and length may reflect specific journal guidelines, evolving reporting guidance should ensure that all standard structural components of a systematic review abstract (ie, background, methods, results, and conclusions) are included.

Critical appraisal tools: With 121 different tools published, “there is no ‘gold standard’ for any study design, nor is there any widely accepted generic tool that can be applied equally well across study types” [35]. This issue emphasizes the importance of selecting the most appropriate available tools and the need for more standardized, pragmatic tools that align with the objectives of systematic reviews on DHI effects.

The appraisal process may demand an assessment or validation and consolidation or adaptation of available tools to ensure it is capturing pertinent research without unnecessarily excluding valuable studies due to overly stringent or irrelevant criteria of appraisal tools.

Evidence synthesis: Incorporating alternative synthesis methods beyond meta-analysis is recommended. Meta-analysis is often challenged by substantial heterogeneity and variability in research questions, study designs, interventions, and outcomes, and incomparable data from the primary research studies. Recognizing the challenges of meta-analyses, systematic reviews should more often explore thematic analyses and other quantitative aggregations to ensure a comprehensive evidence base. Purely narrative approaches or summary tables of primary studies lack the quantitative aggregation needed for conclusive reviews.

Conclusion validity: Researchers should strive for outcome measures that are sensitive to change and aim to capture both relevant proximal and distal intervention effects to ensure meaningful and legitimate conclusions. In interpreting findings, conclusions that suggest limited evidence of benefits due to outcomes comparable with standard or conventional care may neglect the significance of parity when additional benefits, such as cost reduction, improved timeliness, and enhanced access to care, are considered. This is particularly pertinent to telemedicine, telerehabilitation, and telehealth.

Limitations

This study has several limitations that should be considered when interpreting the findings.

First, all data were extracted exclusively from abstracts. We acknowledge that abstracts may not fully capture the methodological details of the full texts; our analysis, therefore, reflects reported, not necessarily executed, methodological rigor. However, abstracts are the primary entry point for study screening and selection, making them critical for evidence recognition.

Second, the analysis was descriptive and exploratory in nature. No causal claims can be made, and the moderate correlations observed should be interpreted as indicative rather than predictive.

Third, data extraction and analysis were based on a random sample rather than a full systematic review of all eligible studies. Thus, while the sample was randomly drawn and diverse in terms of publication outlets and topics, it may not fully represent the entire body of systematic reviews on DHIs.

Finally, although a structured coding scheme was applied, some degree of subjective judgment was inevitable in the classification of PICO elements and evidence conclusiveness.

Future studies should complement abstract-level analyses with full-text reviews, interrater reliability testing, and more refined models to account for complex interactions between review components.

Conclusions

This sample-based meta-research study provides empirical insight into why systematic reviews on DHIs often fail to yield conclusive evidence.

The findings underscore the crucial role of optimally specifying the PICO elements of interest in DHI research, revealing that inadequate PICO specification correlates somewhat with suboptimal evidence recognition.

XSL•FO
RenderX

However, while clearer specification—especially of outcomes and clinical problems—was moderately associated with stronger evidence, this alone did not guarantee evidentiary clarity.

Our findings suggest that methodological coherence across all review stages is a necessary condition to ensure conclusions and evidence-informed decisions in the digitalization of healthcare are both valid and meaningful.

In summary, beyond question formulation, issues in search strategy, eligibility criteria, and synthesis design also play a critical role in shaping review outcomes. To support the underlying processes, we propose a structured, PICO-based categorization framework that is aligned with this research, builds on well-established categories, and provides clear and differentiated definitions and associated terms. Such a framework could inform the research question, and the search strategy and screening process by providing terms for search strings and inclusion or exclusion criteria. Our proposed framework could ultimately enable comprehensive and continuous evidence and gap mapping for DHIs, especially if the increasing power of artificial intelligence tools is leveraged in the research process. Thus, it may help to improve the quality and transparency of systematic reviews—especially in the complex and dynamic field of digital health.

Acknowledgments

This study would not have been possible without the support of the Health and

Social Informatics Research Group of the University of Applied Sciences, Hochschule Osnabrück. We thank Dr Moritz Esdar for assistance with the development and execution of the search strategy, as well as Jan-Oliver Kutza, Dr Johannes Thye, and Prof Ursula Hübner for their contributions to the search terms and in the screening phase. We are grateful to Dr Johannes Thye for his assistance with proofreading.

Funding

This work was funded by the German Ministry of Health and the Lower Saxony Ministry of Science and Culture Hannover, Germany.

Authors' Contributions

Conceptualization: UB (lead) and JDL (equal)
Data curation: UB
Formal analysis: UB
Investigation - Development and execution of search strategy: UB, JDL
Investigation - Initial screening for inclusion and exclusion criteria: UB, JDL
Methodology: UB (lead), JDL (supporting critical review and testing of classification criteria)
Visualization: UB
Writing – original draft: UB
Writing – review & editing: UB, JDL

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strings by database.
[DOCX File, 38 KB - ojphi_v18i1e78210_app1.docx]

Multimedia Appendix 2
Inclusion and exclusion criteria.
[DOCX File, 28 KB - ojphi_v18i1e78210_app2.docx]

Multimedia Appendix 3
Classification criteria.
[DOCX File, 46 KB - ojphi_v18i1e78210_app3.docx]

Multimedia Appendix 4
Sample, charting, and analyses.
[XLSX File, 783 KB - ojphi_v18i1e78210_app4.xlsx]

XSL•FO
RenderX

Checklist 1
PRISMA-ScR checklist.
[PDF File, 97 KB - ojphi_v18i1e78210_app5.pdf]

References

1. Aagja J, Shome S, Chandra A. A bibliometric analysis of digital health & mobile health related global research publications. Hosp Top 2023;101(4):319-325. [doi: 10.1080/00185868.2022.2060155] [Medline: 35380102]
2. Babić A, Poklepović Peričić T, Pieper D, Puljak L. When is the evidence conclusive? Analysis of systematic reviews for which Cochrane declared that conclusions will not change with further studies. Res Synth Methods 2022 Jul;13(4):478-488. [doi: 10.1002/jrsm.1556] [Medline: 35278034]
3. Kang T, Zou S, Weng C. Pretraining to recognize PICO elements from randomized controlled trial literature. Stud Health Technol Inform 2019 Aug 21;264:188-192. [doi: 10.3233/SHTI190209] [Medline: 31437911]
4. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123(3):A12-A13. [Medline: 7582737]
5. Cañón M, Buitrago-Gómez Q. The research question in clinical practice: a guideline for its formulation. Revista Colombiana de Psiquiatría (English ed) 2018 Jul;47(3):193-200. [doi: 10.1016/j.rcpeng.2016.06.003]
6. Hosseini MS, Jahanshahlou F, Akbarzadeh MA, Zarei M, Vaez-Gharamaleki Y. Formulating research questions for evidence-based studies. J Med Surg Public Health 2024 Apr;2:100046. [doi: 10.1016/j.glmedi.2023.100046]
7. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. AMIA Annu Symp Proc 2006;2006:359-363. [Medline: 17238363]
8. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7(1):16. [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
9. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
10. Andrade C. The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. Indian J Psychol Med 2019;41(3):210-215. [doi: 10.4103/IJPSYM.IJPSYM_193_19] [Medline: 31142921]
11. Kuckartz U, Rädiker S, Ebert T, Schehl J. Statistik: Eine Verständliche Einführung 2, Überarbeitete Auflage: Springer VS; 2013.
12. Session three: a “snapshot” of the steps of conducting a Cochrane Review (part 1). YouTube. 2011 Aug 24. URL: https://www.youtube.com/watch?v=HaW3ew-ezPk [accessed 2024-12-12]
13. Higgins JPT, Thomas J, Chandler J, et al. Cochrane handbook for systematic reviews of interventions version. In: Part 2: Core Methods 2024. URL: https://training.cochrane.org/handbook/current/part-2 [accessed 2024-12-12]
14. Blum D, Raj SX, Oberholzer R, et al. Computer-based clinical decision support systems and patient-reported outcomes: a systematic review. Patient 2015 Oct;8(5):397-409. [doi: 10.1007/s40271-014-0100-1] [Medline: 25432150]
15. Buford A, Ashworth HC, Ezzeddine FL, et al. Systematic review of electronic health records to manage chronic conditions among displaced populations. BMJ Open 2022 Sep 6;12(9):e056987. [doi: 10.1136/bmjopen-2021-056987] [Medline: 36285578]
16. Rostam Niakan Kalhori S, Rahmani Katigari M, Talebi Azadboni T, Pahlevanynejad S, Hosseini Eshpala R. The effect of m-health applications on self-care improvement in older adults: a systematic review. Inform Health Soc Care 2023 Jul 3;48(3):292-331. [doi: 10.1080/17538157.2023.2171878] [Medline: 36867051]
17. Smith C, Gold J, Ngo TD, Sumpter C, Free C. Mobile phone-based interventions for improving contraception use. Cochrane Database Syst Rev 2015 Jun 26;2015(6):CD011159. [doi: 10.1002/14651858.CD011159.pub2] [Medline: 26115146]
18. Jones OT, Calanzani N, Saji S, et al. Artificial intelligence techniques that may be applied to primary care data to facilitate earlier diagnosis of cancer: systematic review. J Med Internet Res 2021 Mar 3;23(3):e23483. [doi: 10.2196/23483] [Medline: 33656443]
19. Daryabeygi-Khotbehsara R, Shariful Islam SM, Dunstan D, McVicar J, Abdelrazek M, Maddison R. Smartphone-based interventions to reduce sedentary behavior and promote physical activity using integrated dynamic models: systematic review. J Med Internet Res 2021 Sep 13;23(9):e26315. [doi: 10.2196/26315] [Medline: 34515637]
20. Aluga D, Nnyanzi LA, King N, Okolie EA, Raby P. Effect of electronic prescribing compared to paper-based (handwritten) prescribing on primary medication adherence in an outpatient setting: a systematic review. Appl Clin Inform 2021 Aug;12(4):845-855. [doi: 10.1055/s-0041-1735182] [Medline: 34433219]
21. Veldhuis LI, Woittiez NJC, Nanayakkara PWB, Ludikhuize J. Artificial intelligence for the prediction of in-hospital clinical deterioration: a systematic review. Crit Care Explor 2022 Sep;4(9):e0744. [doi: 10.1097/CCE.0000000000000744] [Medline: 36046062]
22. Akazawa M, Hashimoto K. Artificial intelligence in gynecologic cancers: current status and future challenges - a systematic review. Artif Intell Med 2021 Oct;120:102164. [doi: 10.1016/j.artmed.2021.102164] [Medline: 34629152]

XSL•FO
RenderX

23. Sondaal SFV, Browne JL, Amoakoh-Coleman M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. In: Li D, editor. PLOS ONE 2016;11(5):e0154664. [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
24. McAlpine H, Joubert L, Martin-Sanchez F, Merolli M, Drummond KJ. A systematic review of types and efficacy of online interventions for cancer patients. Patient Educ Couns 2015 Mar;98(3):283-295. [doi: 10.1016/j.pec.2014.11.002] [Medline: 25535016]
25. Paalimäki-Paakki K, Virtanen M, Henner A, Nieminen MT, Kääriäinen M. Effectiveness of digital counseling environments on anxiety, depression, and adherence to treatment among patients who are chronically ill: systematic review. J Med Internet Res 2022 Jan 6;24(1):e30077. [doi: 10.2196/30077] [Medline: 34989681]
26. Bul KCM, Bannon C, Krishnan N, Dunlop A, Szczepura A. Can eHealth applications improve renal transplant outcomes for adolescents and young adults? A systematic review. Transplant Rev (Orlando) 2023 Apr;37(2):100760. [doi: 10.1016/j.trre.2023.100760] [Medline: 37172527]
27. Edison MA, Connor MJ, Miah S, et al. Understanding virtual urology clinics: a systematic review. BJU Int 2020 Nov;126(5):536-546. [doi: 10.1111/bju.15125] [Medline: 32463991]
28. Mokaya M, Kyallo F, Vangoitsenhoven R, Matthys C. Clinical and patient-centered implementation outcomes of mHealth interventions for type 2 diabetes in low-and-middle income countries: a systematic review. Int J Behav Nutr Phys Act 2022 Jan 6;19(1):1. [doi: 10.1186/s12966-021-01238-0] [Medline: 34991593]
29. Shen H, van der Kleij R, van der Boog PJM, Chang X, Chavannes NH. Electronic health self-management interventions for patients with chronic kidney disease: systematic review of quantitative and qualitative evidence. J Med Internet Res 2019 Nov 5;21(11):e12384. [doi: 10.2196/12384] [Medline: 31687937]
30. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. Heart Lung Vessel 2013;5(4):219-225. [Medline: 24364016]
31. Piantadosi S. Clinical Trials: A Methodologic Perspective, 1st edition: Wiley; 2005. [doi: 10.1002/0471740136]
32. Hrynyschyn R, Prediger C, Stock C, Helmer SM. Evaluation methods applied to digital health interventions: what is being used beyond randomised controlled trials?-A scoping review. Int J Environ Res Public Health 2022 Apr 25;19(9):5221. [doi: 10.3390/ijerph19095221] [Medline: 35564616]
33. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020;3(1):110. [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
34. Ahuja AS. Should RCT's be used as the gold standard for evidence based medicine? Integr Med Res 2019 Mar;8(1):31-32. [doi: 10.1016/j.imr.2019.01.001] [Medline: 30805294]
35. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VS, Grimmer KA. A systematic review of the content of critical appraisal tools. BMC Med Res Methodol 2004 Sep 16;4(1):22. [doi: 10.1186/1471-2288-4-22] [Medline: 15369598]

Abbreviations

DHI: digital health intervention
PICO: population, problem, or patient, intervention, comparison, and outcome
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial
RoB: risk of bias

Edited by E Mensah; submitted 28.May.2025; peer-reviewed by A Babuci, C Obianyo, N Levitz; revised version received 11.Dec.2025; accepted 17.Jan.2026; published 16.Apr.2026.

Please cite as:
Buddrus U, Liebe JD
Evaluating Methodological Coherence and Evidence Recognition in Digital Health Systematic Reviews: Sample-based Meta-research Study
Online J Public Health Inform 2026;18:e78210
URL: https://ojphi.jmir.org/2026/1/e78210
doi:10.2196/78210

© Uwe Buddrus, Jan-David Liebe. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 16.Apr.2026. 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,

XSL•FO
RenderX

provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Patient Portals on Asthma and Chronic Obstructive Pulmonary Disease–Related Quality-of-Care Outcomes: Systematic Review

Sukriti Kc[1], MPH; Nicholas S Hopkinson[2], PhD; Ana Luisa Neves[3], PhD; Azeem Majeed[1], MD; Anthony A Laverty[1], PhD

[1]Department of Primary Care and Public Health, Imperial College London, 90 Wood Lane, London, United Kingdom

[2]National Heart and Lung Institute, Imperial College London, London, United Kingdom

[3]Global Digital Health Unit, Imperial College London, London, United Kingdom

Corresponding Author:
Sukriti Kc, MPH
Department of Primary Care and Public Health, Imperial College London, 90 Wood Lane, London, United Kingdom

Abstract

Background: Patient portals are increasingly used to support self-management and facilitate care coordination in people with chronic diseases. Although some asthma quality-of-care gains and modest chronic obstructive pulmonary disease (COPD) improvements have been reported in wider digital health studies, a comprehensive review of patient portal-specific evidence is lacking.

Objective: This study aimed to systematically synthesize evidence on the impacts of patient portals on adult asthma and COPD outcomes across 6 quality-of-care domains (effectiveness, patient-centeredness, equity, efficiency, safety, and timeliness).

Methods: We searched 5 databases (MEDLINE/PubMed, Embase, PsycINFO, CINAHL, and Google Scholar) for quantitative evaluations published up to March 2026 (CRD42022316044). The Newcastle-Ottawa Scale and Cochrane Risk of Bias-2 tool were used for quality assessments, and certainty of evidence was evaluated using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach. Findings were narratively synthesized using the SWiM (Synthesis Without Meta-analysis) framework.

Results: We identified 4038 records, and 16 papers published between 2007 and 2024 met the inclusion criteria (asthma=13, COPD=2, both=1). Three were randomized controlled trials, all rated high risk of bias. Of the other studies, 5 had low risk, 6 had moderate risk, and 2 had high risk of bias. Effectiveness and patient-centeredness were most frequently reported, with 12 papers each. Eight papers were on equity, 5 each on efficiency and safety, and 4 on timeliness. In asthma, modest disease control improvements were reported in 27.8% (42/151) of participants using patient portals with home visit support versus portal use alone (35/150, 23.3%). In COPD, one large, interrupted time series study (N=909,724) reported post-implementation reductions in hospitalizations (slope change from 1.33 to −4.38 per 10,000 patients per month; $P<.001$) but no mortality benefits. Short-term improvements in asthma-related quality of life were observed among patient portal users (mean Mini Asthma Quality of Life Questionnaire score change of 0.67; 95% CI 0.36-0.98) but with no sustained effects. In terms of equity, several digital access barriers were noted. Assessments of efficiency outcomes were largely subjective, and safety events were weakly described. Across all outcome domains, GRADE assessment indicated low to very low evidence certainty.

Conclusions: This review found low to very low certainty of evidence on the impacts of patient portals on quality of care for asthma and COPD. While modest disease control improvements and short-term quality-of-life gains were noted in asthma, and reduced hospitalizations in COPD were observed, the evidence base remains limited and methodologically weak. This review cautions against a reliance on patient portals for clinical management of asthma and COPD and highlights that more methodologically robust primary studies are needed to identify how portal functions may yield clinically meaningful benefits for patients with chronic respiratory disease.

Trial Registration: PROSPERO CRD42022316044; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022316044

(Online J Public Health Inform 2026;18:e83483) doi:10.2196/83483

KEYWORDS

patient portal; asthma; COPD; electronic health record; digital health

XSL•FO
RenderX

Introduction

Chronic respiratory diseases, including asthma and chronic obstructive pulmonary disease (COPD), affect approximately 450 million people worldwide [1]. Asthma prevalence in the United Kingdom ranks among the highest globally, with up to 8 million people living with asthma, and it causes 3 deaths each day [2,3]. Likewise, nearly 3 million people in the United Kingdom live with COPD, and it is one of the leading causes of emergency hospitalizations [4]. Together, these conditions contribute to a significant economic burden on the National Health Service (NHS), with an approximate annual cost of £3 billion (£1=US $1.33 as of July 27, 2026) for asthma and a further £1.9 billion for COPD [5]. Despite the scale of this burden, the quality of respiratory care in the United Kingdom remains suboptimal [6]. A recent national audit highlighted persistent shortfalls in respiratory care across England and Wales, with notable delays in timely care and missed opportunities for early intervention [6]. These gaps underscore the need for approaches that extend beyond episodic clinical encounters and support ongoing disease management.

With the advent of electronic health records (EHRs), digital services such as patient portals are increasingly recognized as tools that support patient engagement in chronic disease care [7]. As services that are linked to provider EHRs, patient portals typically provide access to personal health information, test results, and offer patient-provider communication options, which enable patients to more actively participate in their care processes [7]. These functions are particularly relevant for asthma and COPD, where symptom variability and frequent treatment adjustments are common, and effective management for both conditions relies on sustained patient engagement using shared decision-making strategies [8,9].

Evidence suggests that patient portals can support this role and offer several patient-specific benefits, such as timely access to health information, enhanced patient-provider communication, and improved care satisfaction [10,11]. Portal functionalities that allow remote assessment and real-time symptom checks may also support workflow efficiencies and contribute to more timely clinical review [12]. Among patients with chronic health conditions, patient portal use has been associated with improvements in selected patient engagement indicators, including disease knowledge, treatment adherence, and health care usage in some circumstances [13]. However, a well-documented limitation in the patient portal literature is the substantial variation in portal features, which are often tailored to disease-specific pathways, making it difficult to assume generalized benefits across conditions [14]. This highlights the need for a more focused evaluation in asthma and COPD to understand how patient portals influence the multidimensional needs of people living with these chronic respiratory diseases.

The Institute of Medicine provides one such framework for a structured evaluation of care quality across domains of effectiveness, safety, patient-centeredness, efficiency, timeliness, and equity impact [15-17]. Effectiveness relates to the delivery of evidence-based care and health outcome improvements, while safety refers to safeguards against errors and adverse events. Likewise, patient-centeredness is based on approaches that address individual needs, whereas efficiency relates to the avoidance of waste in the use of economic, human, and technical resources. This framework also considers timeliness, which refers to reductions in delays to care, and equity impacts, which relate to improvements in avoidable disparities in service provision [15].

Findings across care quality indicators for asthma and COPD remain fragmented and often do not separate patient portals from other technologies, or they are focused on a limited subset of outcomes. For example, a recent systematic review noted patient portal benefits related to some markers of effective disease management, such as medication adherence and reduced emergency visits, but this evidence was restricted to the pediatric population with asthma [18]. Another review identified some efficiency benefits, such as enhanced access for asthma care, but did not distinguish the specific contributions of patient portals from those of wider digital tools [19]. Similar improvements related to treatment adherence, unscheduled health care service use, and asthma control were also noted in a recent Cochrane review, but this study also did not isolate the effects of patient portals from other interventions [20]. A recent review further identified some patient portal-related improvements in exacerbation self-management and quality of life (QoL) in COPD, but the overall results remained inconclusive [21]. As no review currently provides a comprehensive overview of the care quality outcomes associated with patient portals in asthma and COPD, an updated evidence synthesis is required. Therefore, this systematic review aimed to provide up-to-date evidence on the impacts of patient portals on asthma and COPD, with a primary objective of evaluating effects across the 6 domains of quality of care.

Methods

Eligibility Criteria

We considered all primary research using quantitative study designs that included adults (aged 17 years and older) with asthma and/or COPD at any disease stage using patient portals. This age cutoff is in line with UK clinical practice guidelines for the management of respiratory conditions such as asthma in adults [22]. As there is no single definition of a patient portal, we used a pragmatic approach to define patient portals as provider EHR-linked tools that offer patients access to their personal health information and enable care coordination using functions such as record views, prescription requests, communication options, appointment booking, and so on [23-25]. To ensure the validity of the reported outcome measures, services that offer portal functions but lack health record integration (ie, stand-alone symptom trackers, medication reminder tools, symptom alert systems, etc) were excluded. Comparators included usual care or other forms of self-management, and studies were included if they reported on at least one of the 6 quality-of-care domains. Mixed methods studies were eligible for inclusion, but only quantitative outcome data were considered. The full eligibility criteria are detailed in Textbox 1.

XSL•FO
RenderX

Textbox 1. Study inclusion and exclusion criteria used for the screening decision to identify all eligible studies.

Inclusion criteria

- All quantitative study designs, including randomized controlled trials, cluster randomized trials, quasi-experimental, case-control, cohort studies, and so on.
- Studies including adults aged 17 years and older with a diagnosis of asthma or chronic obstructive pulmonary disease.
- Studies in English.
- All available studies published to date.

Exclusion criteria

- Studies that do not include primary empirical data.
- Services that are not linked to electronic health records, or those that provide non–electronic health record–based services such as appointment or instruction reminders, telemonitoring tools, and online educational platforms, as well as applications that rely on self-monitoring and self-recording of symptoms.
- Full text not available.

Search Strategy

The review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidance and the SWiM (Synthesis Without Meta-analysis) reporting guideline [26,27]. The study protocol was registered with PROSPERO (International Prospective Register of Systematic Reviews) (CRD42022316044). The search strategy was developed and reported according to the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) guidelines [28]. Electronic searches of key databases such as MEDLINE(R), Embase, PsycINFO via OVID, CINAHL via EBSCOhost, and Google Scholar were conducted in January 2025 and updated in March 2026 using keywords and the MeSH terms from the National Library of Medicine. The full electronic search strategies for all databases, including complete Boolean search strings and limits used, are detailed in “Full Search Strategy for All Databases” in Multimedia Appendix 1 [29-44].

Bibliographies of included studies were also screened to identify additional papers not captured in the primary search. Attempts to find missing data and access to studies that potentially met the inclusion criteria but lacked full texts were made by contacting the corresponding authors and source journals. Analyses were conducted using the available data.

Selection Process

Two reviewers (SK and AAL) independently conducted initial title and abstract screening, followed by full-text evaluation to determine study eligibility for inclusion. Covidence software was used for deduplication and screening, and all conflicts were resolved by discussion. Data extraction and quality assessment were conducted by 1 reviewer (SK) and checked by a second reviewer (AAL). Data were extracted using a standardized computer-based spreadsheet and included details of the study, population, intervention, comparator, and outcomes. Only quantitative data were extracted from mixed methods studies.

Data Items and Synthesis

Outcomes were categorized using the Institute of Medicine (IOM) quality-of-care guide for evidence synthesis. The IOM framework offers a structured approach to understand the quality of healthcare services, and it has been previously used to assess digital health interventions [15,45]. Extracted data categorized according to the IOM framework included indicators of 1. Effectiveness (eg, disease and symptom management, medication use, activity limitation, lung function, etc), 2. Patient-centeredness (eg, indicators of self-management, knowledge, QoL, satisfaction, etc), 3. Equity impacts (eg, socioeconomic outcomes or other determinants of health), 4. Efficiency (eg, economic outputs, streamlined administrative tasks, etc), 5. Safety related to both patient health and technology (eg, adverse events, privacy, security, etc) and 6. Timeliness (eg, time to care, information exchange timeliness, etc).

This review followed the SWiM reporting guideline to ensure transparency in narrative synthesis [26]. Studies were grouped according to (1) disease type (asthma or COPD) and (2) the IOM quality-of-care domains (effectiveness, patient-centeredness, equity impacts, efficiency, safety, and timeliness). A narrative synthesis of results was performed as the studies had different designs, differing availability of and exposure to patient portal functionalities, and different outcomes assessed using a range of instruments. Additionally, there were few studies in each exposure-outcome pair and so, in line with other studies [18,46] and guidance [47], we did not consider that meta-analysis would be informative. Due to the limited number of studies per outcome, formal assessment of publication bias using funnel plots was also not performed and this was assessed narratively. Reported effect measures included mean differences, proportions, regression coefficients, and estimates of trends, depending on the study design and outcome type.

Outcomes were standardized by category with effects reported as positive, no effect, or negative within each domain in the visual representation using a harvest plot. A structured narrative synthesis was undertaken, supported by tabulation of study characteristics (Multimedia Appendix 2) and domain-level summaries (). In reporting, greater emphasis was placed on findings from studies with a lower risk of bias and larger sample sizes. Variability in results was explored in relation to differences in study design, population characteristics, portal functionalities, and follow-up duration.

XSL•FO
RenderX

Risk-of-Bias Assessment

Quality assessment was using the original Newcastle-Ottawa Scale for cohort studies and an adapted version from a previous review for cross-sectional studies [48,49]. For randomized controlled trials (RCTs), the Cochrane Risk of Bias Tool 2.0 was used [50]. This was a deviation from the registered study protocol (PROSPERO:CRD42022316044). Although the protocol prespecified the use of the Joanna Briggs Institute Critical Appraisal Tool, the Newcastle-Ottawa Scale and the Cochrane Risk of Bias Tool 2.0 were ultimately used as they supported study design–specific assessment and domain-level evaluation, allowing for a more transparent and rigorous appraisal of bias. Quality assessments were carried out independently by 1 reviewer (SK) and verified by a second (AAL). The quality of evidence across each of the IOM domains was also assessed using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach to support the strength of the recommendation [51,52]. Using GRADE, the narrative summary of effect for each outcome domain was assessed by considering the risk of bias across studies, their indirectness, imprecision, inconsistency, and publication bias [51,52].

Results

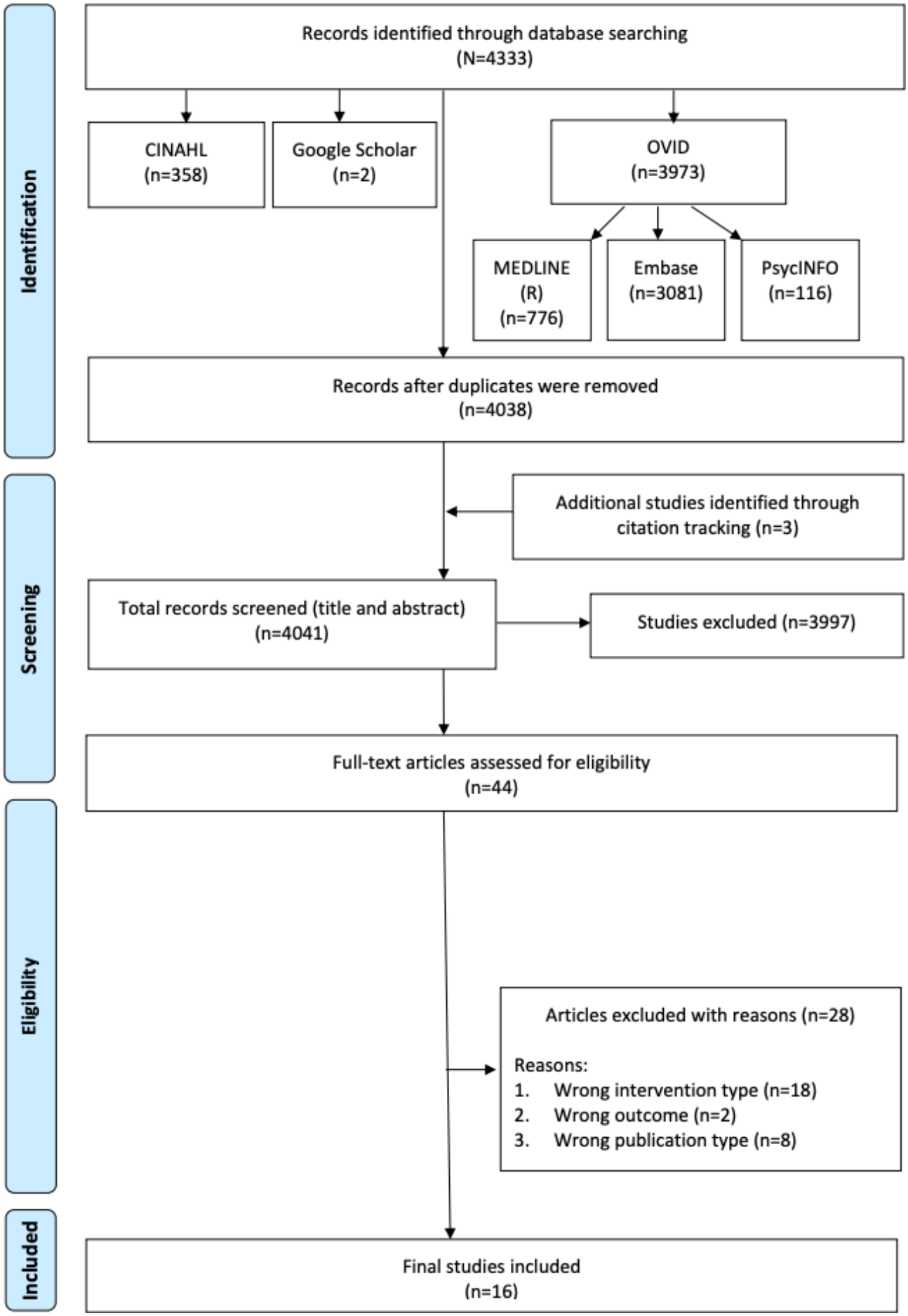
Study Selection

The search identified 4038 citations after removing duplicates (Figure 1), which were retained for abstract screening, and 3 additional papers were identified through backward or forward screening. We also identified 4 studies at the title and abstract screening stage that met our inclusion criteria, but despite our best attempts to contact the authors, full texts were not available.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing the database search detailing the sources used, number of records identified per source, reasons for exclusion, and the number of full-text reports assessed for the final inclusion decision. CINAHL: Current Index to Nursing and Allied Health Literature.

XSL•FO
RenderX



XSL•FO
RenderX

A total of 3998 papers were excluded due to a lack of relevance to asthma or COPD, or the absence of any digital health component. Subsequently, 44 papers were shortlisted for full-text screening, of which 28 were excluded (18 studies included wrong interventions and stand-alone tools, 8 were the wrong publication type, and 2 reported irrelevant outcomes). Of the included studies, some of the publications were from the same author group, with 3 papers linked to a single trial [32,33,35] and 2 to another RCT [39,42]. Although 1 pilot study [36] was related to a subsequent cohort analysis [37], trial registration numbers and population samples were different, so they were considered as separate studies. Therefore, a total of 16 papers were included in the review, which represented 14 unique studies [29-44].

Study Characteristics

Of the 16 included papers, 13 investigated asthma only and 2 COPD only [38,41]. One study included both asthma and COPD populations and did not separate these 2, or report individual disease-specific outcomes [43]. Publication years ranged from 2007 to 2024. Eight papers were from the United States [29,31-33,35-37,40], 3 from Canada [30,39,42], 2 from the Netherlands [43,44], and 1 each from Portugal [34], Spain [38], and the United Kingdom [41]. There were 3 RCTs [30,32,34], 5 observational studies [31,33,35,39,41], and 4 quasi-experimental studies [29,38,42,44]. Two were feasibility evaluations [36,40]. One was a cohort trial [37] and another was a cross-sectional study [43]. Details of the reports included are shown in Multimedia Appendix 2.

Risk-of-Bias in Studies

In the risk-of-bias quality assessment, 5 out of 16 were rated as high, 6 moderate, and 5 low risk of bias. All of the RCTs received a high risk-of-bias rating due to high attrition and missing outcome data [30,32,34]. Of the other studies, 5 had low risk of bias [1,33,35,38,41], 6 had moderate risk [36,37,39,40,43,44], and 2 had high risk of bias [29,31], mainly due to the absence of a comparator group and a lack of treatment effect assessment against a baseline. Figure 2 details the methodological quality and risk-of-bias decisions for all included papers.

Figure 2. Risk-of-bias rating for the included studies [29-44]. The figure illustrates the risk-of-bias assessment of the included studies with details of the article name, publication year, categories of assessment, and the overall risk-rating decision where high indicates poor quality papers with high risk of bias, moderate indicates medium risk of bias, and low indicates good-quality papers with low risk of bias. The Newcastle-Ottawa Scale (NOS) was used for cohort studies, and an adapted version of the NOS tool was used for cross-sectional studies [48,49]. For NOS screening decisions, a total score of <4 indicates high (red), 5 - 6 indicates moderate (yellow), and >7 indicates low risk of bias (green). The Cochrane Risk of Bias Tool 2.0 was used for randomized controlled trials where the overall rating was "high" (red) if any domain was rated high, "low" (green) only if all domains were low, and otherwise "unclear" [50].

Cohort/ Case control studies

Author (Year)	Selection (Stars)	Comparability (Stars)	Outcome/Exposure (Stars)	Total	Overall risk rating
Acharya and Sarraf (2017)	2/4	0/2	2/3	4	High
Alladina et al (2024)	2/4	0/2	2/3	4	High
Apter et al (2020)	4/4	2/2	3/3	9	Low
Localio et al (2022)	4/4	1/2	3/3	8	Low
Mammen et al (2021)	3/4	0/2	3/3	6	Moderate
Mammen et al (2022)	2/4	0/2	3/3	5	Moderate
Martínez Nicolás et al (2019)	4/4	1/2	3/3	8	Low
Morita et al (2019)	3/4	0/2	3/3	6	Moderate
Rudin et al (2019)	2/4	0/2	3/3	5	Moderate
Taylor et al (2023)	4/4	1/2	3/3	8	Low
To et al (2020)	3/4	2/2	3/3	8	Low
Tossaint-Schoenmakers et al (2021)	2/4	1/2	2/2	5	Moderate
Zijp et al (2024)	3/4	0/2	3/3	6	Moderate

Randomized controlled trials

Author (Year)	D1: Randomization	D2: Deviations from Interventions	D3: Missing Data	D4: Outcome Measurement	D5: Reporting	Overall risk rating
Ahmed et al (2016)	Low	Some concerns	High	High	Some concerns	High
Apter et al (2019)	Low	Some concerns	High	Low	Low	High
Cuz-Correia et al (2007)	Low	Some concerns	High	High	Some concerns	High

When using GRADE to rate the certainty of domain-level evidence using a narrative summary [52], certainty was judged low to very low for both asthma and COPD, reflecting serious limitations in study design, inconsistency, imprecision, and indirectness (Table 1). Narrative inspection and GRADE assessment suggested variation across studies in terms of their design, population characteristics, intervention scope, and implementation, which may partly account for the inconsistencies in outcomes reported and the differences in findings. Where risk of bias from missing results was suspected, checks of partially published data and related publications were conducted, and authors were contacted to request additional data. Publication bias was not deemed a concern in asthma as studies generally varied in their size, design, geographical representation, and direction of effect, with no consistent pattern of positive findings. In COPD, the findings across all domains were drawn from a limited number of papers, which was considered a publication bias risk. There was also a risk of sample overlap in some of the studies, which was considered when grading the certainty of evidence. For example, Apter et

XSL•FO
RenderX

al [32,33] and Localio et al [35] related to the same trial and consisted of the same participants, which may lead to overestimations in the true sample size. The risk of overlap was less clear for Mammen et al [36,37], as the total sample size and group composition (eg, age range) differ, and the authors do not clearly describe whether the participants in the larger cohort trial included the smaller β-test group.

Table . Certainty of evidence across the quality-of-care outcome domains assessed using the GRADE[a] approach.

Domains	Studies, n	Report type	Sample size	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall GRADE level of evidence
Asthma[b]									
Effectiveness	10	3 RCTs[c], 7 non-RCTs	1733 (risk of sample overlap)	Serious concerns	Serious concerns	Not serious	Very serious	Not serious	Very low
Patient-centeredness	11	2 RCTs, 9 non-RCTs	925 (risk of sample overlap)	Serious concerns	Serious concerns	Not serious	Very serious	Not serious	Very low
Efficiency	5	1 RCT, 4 non-RCTs	238 (risk of sample overlap)	Serious concerns	Not serious	Very serious	Very serious	Not serious	Very low
Equity impacts	7	1 RCT, 6 non-RCTs	1108 (risk of sample overlap)	Not serious	Not serious	Serious concerns	Serious concerns	Not serious	Low
Safety	5	1 RCT, 4 non-RCTs	491	Not serious	Serious concerns	Serious concerns	Serious concerns	Not serious	Very low
Timeliness	3	1 RCT, 2 non-RCTs	63 (risk of sample overlap)	Serious concerns	Not serious	Serious concerns	Serious concerns	Not serious	Very low
Chronic obstructive pulmonary disease[d]									
Effectiveness	2	2 non-RCTs	910,222	Serious concerns	Not serious	Not serious	Not serious	Serious concerns	Very low
Patient-centeredness	2	2 non-RCTs	1246	Serious concerns	Serious concerns	Not serious	Serious concerns	Serious concerns	Very low
Equity impacts	2	2 non-RCTs	1246	Serious concerns	Serious concerns	Not serious	Serious concerns	Serious concerns	Very low
Timeliness	1	1 non-RCT	498	Serious concerns	Not serious	Not serious	Serious concerns	Serious concerns	Very low

[a]GRADE: Grading of Recommendations Assessment, Development, and Evaluation.

[b]The table summarizes the GRADE assessment for certainty of evidence for asthma. “Domains” refer to the 6 care quality indicators for which pooled effect estimates were not available and only a narrative synthesis of the evidence was provided. Evidence from randomized controlled trials was initially rated as high certainty and observational studies as low certainty, with downgrading applied for risk of bias, inconsistency, indirectness, imprecision, and publication bias, where appropriate. Certainty ratings were categorized as high, moderate, low, or very low, reflecting confidence that the estimated effect is close to the true effect. Concerns were reported as not serious, serious concerns (−1), or very serious concerns (−2). The evidence from the study that reported mixed population (ie, asthma and COPD) without separate outcome measurement was considered applicable to both conditions but with downgrading for indirectness. Where risk of sample overlap has been noted, it relates to studies that potentially used the same sample population for the analyses.

[c]RCTs: randomized controlled trials.

[d]Table summarizes the GRADE assessment for certainty of evidence for COPD. “Domains” refer to the 6 care quality indicators for which pooled effect estimates were not available and only a narrative synthesis of the evidence was provided. Evidence from randomized controlled trials was initially rated as high certainty and observational studies as low certainty, with downgrading applied for risk of bias, inconsistency, indirectness, imprecision, and publication bias, where appropriate. Certainty ratings were categorized as high, moderate, low, or very low, reflecting confidence that the estimated effect is close to the true effect. Concerns were reported as not serious, serious concerns (−1), or very serious concerns (−2). The evidence from the study that reported mixed population (ie, asthma and COPD) without separate outcome measurement was considered applicable to both conditions but with downgrading for indirectness. Where risk of sample overlap has been noted, it relates to studies that potentially used the same sample population for the analyses.

XSL•FO
RenderX

Participants

Study participants were recruited from a range of settings. Five publications were multicenter collaborations involving primary care practices, specialty clinics, pulmonary or ENT departments, and/or academic hospitals [31-33,35,39], and 3 papers were each based in primary care [36,37,42], secondary care [38,41,44], and tertiary care centers [29,30,34]. One study was community-based [43] and 1 ambulatory care–based [40]. The sample size ranged from 7 to 9,09,724 participants [36,38]. Participants were predominantly female (55%-90%) [32,38]. Nine publications reported data on ethnicity [31,32,36-40,42], and the number of White participants enrolled ranged from 0.7% to 78.3% [31,32], whereas the number of Black/African American participants ranged from 9.2% to 76.8% [31,32].

Patient Portal Functions

Communication options were included in 11 papers using features such as secure messaging options, email alerts, provider logs, emergency contacts or services, and so on [29-31,34,36-39,41,42,44]. Functions supporting health and disease monitoring were reported in 11 papers using asthma diaries, logs, journals, immunization records, wireless spirometry, and so on [29,31-34,36,37,39,42,44]. Self-management support and educational components using personalized action plans, social forums, training, and linked resources were mentioned in 9 papers [29-31,34,36,37,41,42,44]. Seven papers included provider-facing dashboards or decision support tools [30,31,34,36,37,41,44]. Appointment management options (eg, review, scheduling, and reminders) were reported in 4 papers [32,33,35,38], and 5 papers offered visual care indicators using features such as disease trend charts, traffic light displays, and so on [31,34,39,42,43]. Three papers included medication lists or review and repeat prescription order functions [32,33,35]. Medication management tools using inhaler logs and trackers appeared in 2 papers [29,31]. One study also offered options to customize monitoring schedules via a clinician dashboard [34]. Six out of the 16 papers offered access to provider records, which mostly consisted of structured data such as laboratory results, action plan updates, clinical decision support outputs, and so on [32,33,36-39,42,43]. Data on quality outcomes related to specific functions were not examined in sufficient detail by any of the included studies to support meaningful conclusions regarding the effects of individual features.

Comparators

Comparators were patients receiving regular care [30,42], portal access but without home visits [32,33], paper diaries and other web-based tools or educational material [34], nonusers of the intervention [41], and English versus Spanish-language speakers [35].

Outcomes

There were differing numbers of publications across different quality domains: effectiveness (12 papers), patient-centeredness (12 papers), equity (8 papers), efficiency (5 papers), safety (5 papers), and timeliness (4 papers). Details of the outcomes for the individual studies are presented in Table 2, and Figure 3 shows a harvest plot with intervention effects in accordance with the SWiM guidance [26].

XSL•FO
RenderX

Table . Summary of reported outcomes categorized according to the 6 quality-of-care domains. The table provides a summary of the outcomes categorized according to the 6 quality-of-care domains of effectiveness, patient-centeredness, efficiency, equity, safety, and timeliness among adults with asthma and chronic obstructive pulmonary disease using patient portals. Could all RCTs in the table have the superscript **c**, to make clear they all refer to "Randomised Controlled Trial"?

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
Acharya and Sarraf (2017) [29]	• Health care service use: 83% of health care providers reported that the use of the Pro-Care application reduced the number of office visits for patients by over 60%.	• Self-management: The majority of the patients strongly agreed that the real-time feedback was very useful in managing their asthma condition. Most patients found the application helpful in keeping track of their prescribed action plans. • Satisfaction: The majority of the patients found the application to be successful in asthma management.	Not stated	Not stated	• Security: The IT group reported no instances of instability or unavailability during the evaluation period.	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
Ahmed et al (2016) [30]	• Disease control: No significant between-group differences over time in asthma control status.	• QoL[a]: Between baseline and 3 months, only the intervention group showed a significant improvement in MAQLQ[b] score (mean change 0.67, 95% CI 0.36-0.98). No significant differences between 3 and 6 months (mean change −0.01, 95% CI −0.35 to 0.32) and 6 - 9 months (mean change −0.12, 95% CI −0.46 to 0.22) for the intervention group. Also, no significant differences over time for the control group and between 2 groups. • MAQLQ significantly associated with improvement in depression PHQ[c] (mean change −0.27, 95% CI −0.37 to −0.18 for 5 unit change), self-efficacy (mean change 0.24, 95% CI 0.16-0.32) and ACT[d] score (mean change −0.25, 95% CI −0.30 to −0.20).	Not stated	Not stated	Not stated	Not stated
Alladina et al (2024) [31]				Not stated	Not stated	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
	• Disease control: Asthma status was consistent and reported as "good" in >69% of daily entries. Overall, 14% of patients experienced asthma exacerbations. • Medication use: Overall, 76% of patients reporting "bad" asthma status used quick relief or rescue medication.	• Satisfaction: Overall, 66% of patients were satisfied or very satisfied with the app. Overall, 49% rated the app "excellent," 30% "good." Overall, 71% of HCPs[e] reported the app as "very easy" to implement, and 29% reported that the app "moderately" or "slightly" helped patient management. Recommendation: Overall, 66% would likely or very likely recommend the app. • Engagement: Overall, 47% did not use the asthma plan, 52% skipped appointment scheduling, 57% skipped education, and 65% did not message the clinic.	• Streamline of communication and appointments: 44% felt that the app helped discuss asthma with care professionals; 37% said it made appointments smoother.			
Apter et al (2019) [32]		Not stated	Not stated	Intervention benefit: Spanish as the primary language and trust on the internet and the patient portal for clinical information were possible effect modifiers, but these subgroups did not prove to have different benefits for the portal plus home visit intervention.	Not stated	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
	Disease control: • Intervention: 27.8% (42 patients) achieved asthma control (ACQ[f] ≤1.5) at 12 months. • Control: 23.3% (35 patients) achieved asthma control at 12 months. • Intervention: 25% (37 patients) had a clinically important improvement (0.5-point reduction in ACQ). • Control: No patients had clinically important improvement. Health care service use: • Only hospitalizations showed statistically significant improvements in the intervention group at 12 months (−0.53, 95% CI −1.08 to −0.024).					
Apter et al (2020) [33]	Disease control: No association between portal usage rates and asthma control at 12 months compared with baseline.	• QoL: No association between portal usage rates and asthma quality of life at 12 months compared with baseline. • Satisfaction: 62% had little confidence that the portal could improve communication with their doctor.	Not stated	Portal use: Patients with chronic diseases or past hospitalizations were less likely to use the portal, but it was unrelated to asthma severity. Overall, 50% of potential participants had computer access, but some chose not to use the portal.	Privacy: Overall, 16% expressed concerns about portal confidentiality.	Not stated
Cruz-Correia et al (2007) [34]				Not stated		

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
	Disease control: More participants in the intervention group believed that the platform could improve their asthma control (73% vs 50%), treatment adherence (64% vs 53%), and asthma care (80% vs 69%) in the intervention versus paper diary group.	Satisfaction: • More patients rated the intervention as very useful compared with the paper diary (*P*=.038). Viewing previous data was considered much easier with the Internet tool than with paper (*P*=.038). • Overall, 92% were very interested in using the intervention in the future versus 12% for paper tools (*P*=.002). • Adherence dropped after the second visit, especially after weeks 1 and 4. Paper diary showed higher recorded adherence (*P*<.001), but likely overestimated.	Streamline of tasks: • The median time to complete both internet and paper diaries was similar (3 minutes per entry; *P*=.675). • PEF[g]/FEV_1[h] monitoring took less time (median 2 minutes; *P*=.028 for internet and *P*=.036 for paper).		Safety report: • No safety concerns reported. Authors conclude that the technology was considered safe. • Issues related to internet connection (n=9), user interface (n=5), system errors (n=3), and question interpretation (n=2) were reported.	• Reporting timeline: Paper tools lacked timeliness; users often delayed entries (some admitted bulk-filling days later).
Localio et al (2022) [35]	Not stated	Not stated	Not stated		Not stated	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
				Portal use: • Overall, 42% (n=102) of English speakers did not use the portal vs 56% (n=29) of Spanish speakers (Pearson chi-square=3.84; *P*=.05). In total, 49% (n=32) of low-literacy English speakers did not use the portal vs 69% (n=18) of low-literacy Spanish speakers (Pearson chi-square=10.4; *P*=.02). Overall, 25% of English-speaking participants reported barriers to portal use, compared with 76% of Spanish-speaking respondents. Access and knowledge: • In response to "I would use the internet more if..." Overall, 25% of English speakers (out of 182 respondents) and 26% of Spanish speakers (out of 43 respondents) selected "I knew more/learned how." In total, 25% of English speakers (n=182) and 23% of Spanish speakers selected "I had access"		
Mammen et al (2021) [36]				Not stated	Not stated	

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
	Disease control: • 1.5 (SD 1.02) point mean ACQ improvement in asthma control (CI[i] 0.59 - 2.51; *P*=.007) at 3-month follow-up. Significant improvements were seen in morning symptoms, nighttime wakening, activity limitations, and shortness of breath, with greatest effects on reductions in wheezing with mean 1.87 (SD 1.34) (*P*=.011, CI 0.61-3.10). • 14.86% (SD 19.4) mean increase in $FEV1\%_{pred}$[j] (CI −3.09 to 32.80l;. *P*=.089). Health service use: • Preventive health care usage increased significantly (1.86 visits per year vs 0.29 per year prior, CI 0.67-2.47; *P*=.005). Medication use: • Increased prescriptions for controller medications (9.29 prescriptions per year vs 1.57 prescriptions per year, CI 4.85-10.58; *P*=.001).	• QoL: 1.91 (SD 1.53) point mean improvement (*P*=.016, CI 0.50-3.31) evenly distributed across all domains (symptoms, activity limitations, emotional functioning, and environmental stimuli). • Satisfaction: 93.9%	• Economic evaluation: $186.52 per person. • Nursing time: 44.47 minutes per visit.			Time to care: 100% care provider response to follow-up, and home symptom monitoring facilitated timely care.
Mammen et al (2022) [37]			Economic evaluation: $186.52 per person over the 6-month period.		Technical issues: 93.2% of visits showed no disruptive technical issues.	Time to care: 38% of visits occurred after-hours or on weekends.

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
	Disease control: • 80% participants had uncontrolled asthma at baseline and at 6 months, 80% had well-controlled asthma. Patients with worse asthma control at baseline benefited the most from the intervention, with a strong correlation between baseline ACQ and symptom improvement (*r*=-0.82, *P*<.001). • $FEV1\%_{pred}$ increased by 4.2% Medication use: • Adherence increased from 45.58% to 85.29% (CI 14.79-64.62). Guideline-based care: • HCP adherence to the guideline-based therapy increased from 43.3% at baseline (CI 22.11-64.55) to 86.7%.	• QoL: Improvement in QoL was strongly associated with improved control (*r*=0.80; *P*<.001) but not with $FEV1\%_{pred}$ (*r*=0.087; *P*=.648). • Satisfaction: 95.7% acceptability. 29/30 patients and all providers indicated the intervention worked better than usual care.		• Wider health determinants: Improvements in $FEV1\%_{pred}$ were greatest for smokers (+10.27% vs nonsmokers+0.68%, CI 1.72 - 17.45), males (+11.27% vs females +0.11%, CI 3.62 - 18.72), and those with high-school education or less (+7.94% vs any college education −0.071%, CI 0.20 - 15.82). • No other significant differences in intervention effects based on gender, smoking status, education, race or ethnicity, or presence of comorbid mental illness.		
Martínez Nicolás et al (2019) [38]		Not stated	Not stated	Not stated	Not stated	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
	Health service use: • For the COPD[k] group, there was a significant decline in slope for any hospitalizations per 10,000 patients per month (before EPP[l]=1.33, SE 0.3, *P*<.001 vs after=−4.38 SE 0.66; *P*<.001) and emergency department use (before=2.27, SE 0.45, *P*<.001 vs. after=−5.08, SE 1.26; *P*<.001) and any outpatient service use (before=0.002, SE 0.0002; *P*<.001 vs after=−0.004, SE 0.001; *P*<.001) but no change in 30-day hospital readmissions (before=0.33, SD 0.10; *P*<.05 vs after=−1.07, SD 0.59).					
Morita et al (2019) [39]	Not stated		Not stated	Use: • Age (≥50 years) was associated with higher usage. • 55.2% (76/138) of the participants had a smartphone and reported being comfortable or very comfortable with its use, but post hoc analysis did not find an association between platform usage and having a smartphone.	Not stated	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
		Satisfaction: • 63.8% (74/116) of the patients agreed or strongly agreed that the intervention was helpful in the management of their asthma. Patients agreed or strongly agreed that the intervention was helpful in the management of their asthma. • 65.2% (75/115) of the patients were confident that the intervention was correct when it presented the patient's asthma action plan zone of control. Patients were confident that the intervention was correct when it presented the patient's asthma action plan zone of control. • 49.6% (58/117) of the participants agreed or strongly agreed that they would continue to use the intervention after the study. Participants agreed or strongly agreed that they would continue to use the intervention after the study. • The platform				

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
		scored 71.1 (SD 19.9) on the standardized SUS[m] at 12 months indicating good usability. Engagement: • Declined use within the first 4 weeks. Overall, 67.5% (83/123) of the participants used the platform weekly initially and only 57.7% (71/123) used the platform in week 45.				
Rudin et al (2019) [40]	Not stated	Engagement: • On average, 25% of all questionnaires had results severe enough to qualify for a call back from a nurse; however, 77% declined at least 1 option for call back.	Not stated	Program completion: • Female patients had significantly higher weekly questionnaire completion rates than males (88 vs 73%; *P*=.02). • Patients with a bachelor's degree level or higher completed significantly more questionnaires than those with lower education (94% vs 74%; *P*<.01). • Clinical characteristics, age, and ethnicity were not significantly associated with questionnaire completion.	Not stated	Not stated
Taylor et al (2023) [41]			Not stated		Not stated	

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
	Health service use: • Reduction in annual COPD or respiratory-related admissions (mean change=1.21 days in intervention group vs 0.62 days in control) and bed-occupied days (mean change=8.12 days in intervention group vs 3.38 days in control). Mortality: • Lower 12-month mortality rate in intervention vs control (16.9% vs 24.1%), but this difference was not statistically significant (*P*=.215).	QoL: • Health-related QoL (measured via EQ-5D visual analogue scale) remained broadly stable over the course of the study. Disease burden: • Symptom burden (measured via COPD assessment test) also showed no significant change during the study period. Self-management: • Median self-managed exacerbations were 2 per year for the intervention group with a higher median exacerbation of 4 per year in patients with the highest application usage.		Use: • Overall, 77% of the participants were sustained use. Subgroup analysis of participants resident in more socioeconomically deprived postcode areas revealed equivalent usage to the rest of the cohort.		Time to adverse events: • Median time to first COPD or respiratory–related admission or death was longer in the intervention cohort (335 days vs 155 days; unadjusted hazard ratio=0.740, 95% CI 0.550-0.996; *P*=.047).
To et al (2020) [42]	Health service use: • No significant differences between the Breathe group, non-Breathe group, and population controls in hospitalizations, ED visits, and physician office visits.	Not stated	Not stated		Not stated	Not stated
Tossaint-Schoenmakers et al (2021) [43]	Not stated	Self-efficacy: • Higher education levels were associated with lower self-efficacy in managing health; *P*<.10 (*b*=−3.521, 95% CI 6.469 to −0.572; *P*=.02)	Not stated		Not stated	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
				Use: • Higher age and high education were associated with decreased usability: respectively, *b*=−0.094, 95% CI 1147 to −0.042 (*P*<.001); and *b*=−2.512, 95% CI 4.791 to −0.232 (*P*=.03). • Patients with asthma or COPD scored significantly lower than those with no chronic disease (*b*=−3.630, 95% CI −6.545 to −0.715; *P*=.02).		
Zijp et al (2024) [44]	Disease control: • Asthma control improved postintervention (ACQ-6-score: baseline μ=2.1, SD=1.3; postintervention μ=1.3, SD=1.0; *t*=2.61, *P*=.02, d=0.70) Medication use: • No significant change in rescue medication use (Last wk additional medication: baseline μ=2.0, SD=1.5; postintervention μ=1.4, SD=1.1; *t*=-1.63, *P*=.10, *d*=-0.06)		Economic evaluation: • Indirect costs showed no significant change • iPCQ[q] score: Baseline μ=34.8, SD=54.7; postintervention μ=22.8, SD=53.3; *t*=1.057, *P*=.33, *d*=0.37. • iMCQ[r] score: Baseline μ=4.6, SD=3.7; postintervention μ=2.9, SD=3.0; *t*=1.52, *P*=.15, *d*=0.41.	Not stated	• Adverse events: No adverse events reported by patients or providers.	Not stated

XSL•FO
RenderX

Author (year)	Effectiveness	Patient-centeredness	Efficiency	Equity impact	Safety	Timeliness
		QoL: • No significant changes in asthma-related quality of life trends • RIQMON-10[n] score: μ=22.6, SD=6.9; postintervention μ=18.9, SD=7.7; *t*=1.70, *P*=.11, *d*=0.46. EQ-5D-5L score: • μ=1.8, SD=0.7; postintervention μ=1.6, SD=0.6; *t*=1.75, *P*=.10, *d*=0.47. VAS-QoL[o] score: • Baseline μ=53.5, SD=17.5; postintervention μ=67.4, SD=18.9; *t*=-2.08, *P*=.06, d=0.56. • Self-management: Perceived ability to manage asthma remained stable with no significant change (*P*=.84) (PCAQ[p] score: baseline μ=21.4, SD=2.7; postintervention μ=21.6, SD=5.0; *t*=−0.231, *d*=0.06).				

[a]QoL: quality of life.

[b]MAQLQ: Mini-Asthma Quality of Life Questionnaire.

[c]PHQ: Patient Health Questionnaire.

[d]ACT: asthma control test.

[e]HCPs: health care providers.

[f]ACQ: Asthma Control Questionnaire.

[g]PEF: peak expiratory flow.

[h]FEV_1: forced expiratory volume in 1 second.

[i]CI: confidence interval reported as specified in the original study. Where the confidence level was not stated, no level was assumed.

[j]$FEV1\%_{pred}$: forced expiratory volume in 1 second percentage predicted.

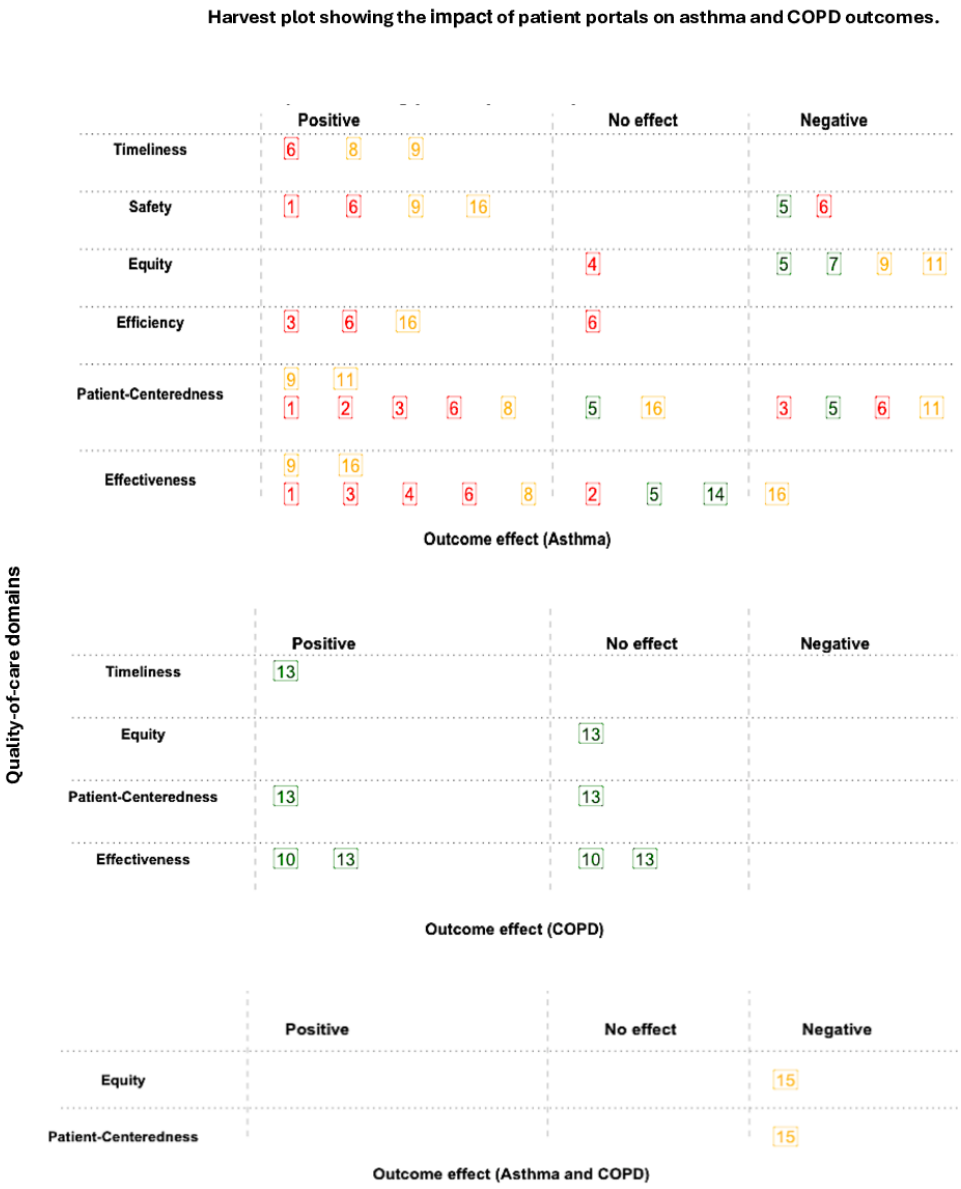
XSL•FO
RenderX

[k]COPD: chronic obstructive pulmonary disease.
[l]EEP: electronic patient portal.
[m]SUS: System Usability Scale.
[n]RIQMON-10: Respiratory Illness Questionnaire-Monitoring 10.
[o]VAS-QoL: Visual Analogue Scale–Quality of Life.
[p]PCAQ: Perceived Control of Asthma Questionnaire.
[q]iPCQ: iMTA Productivity Cost questionnaire.
[r]iMCQ: iMTA Medical Consumption Questionnaire.

XSL•FO
RenderX

Figure 3. The figure illustrates the outcome effects, categorized as "Positive," "No effect," and "Negative" related to the specific domain and their indicators. Multiple outcomes of the same category were counted once, and studies with different outcome results were recorded separately. The box color represents risk-of-bias rating, where red indicates high risk of bias, yellow indicates medium risk of bias, and green indicates low risk of bias. COPD: chronic obstructive pulmonary disease. Studies: (1) Acharya and Sarraf (2017) [29]; (2) Ahmed et al (2016) [30]; (3) Alladina et al (2024) [31]; (4) Apter et al (2019) [32]; (5) Apter et al (2020) [33]; (6) Cruz-Correia et al (2007) [34]; (7) Localio et al (2022) [35]; (8) Mammen et al (2021) [36]; (9) Mammen et al (2022) [37]; (10) Martinez Nicolas et al (2019) [38]; (11) Morita et al (2019) [39]; (12) Rudin et al (2019) [40]; (13) Taylor et al (2023) [41]; (14) To et al (2020) [42]; (15) Tossaint-Schoenmakers et al (2021) [43]; and (16) Zijp et al (2024) [44].

XSL•FO
RenderX



XSL•FO
RenderX

1. Effectiveness

Overall, evidence from the 12 studies on asthma and COPD suggests possible improvements in selected indicators of effectiveness, including asthma control, medication adherence, and health care usage. However, findings were inconsistent and often derived from interventions that included additional support beyond portal use. Evidence for COPD was restricted to 2 observational studies, with reported hospitalization reductions but no mortality benefits. GRADE assessments showed that the certainty of evidence was very low for both asthma and COPD.

Asthma

Seven out of the 10 studies on asthma found evidence of improvements in at least one of the indicators of effectiveness [29,31,32,34,36,37,44]. Four studies showed either no difference in disease control [30,33], health care service use [42], or small, nonsignificant gains in rescue medication use and Asthma Control Questionnaire (ACQ) scores [44]. Among the studies with positive effects, improved disease control was noted in a large RCT rated as high risk of bias, which showed a modest increase in well-controlled asthma (27.8% intervention vs 23.3% control) and clinically significant improvements (0.5-point ACQ reduction in 25% intervention group) at 12 months of follow-up [32]. However, a subsequent paper by the same group found that disease control improvements were unrelated to portal usage rates [33]. In other smaller studies with moderate bias risks, a 1.5-fold improvement in ACQ scores at follow-up was noted (2.1 baseline vs 1.3 postintervention) [44]. There were also improvements in guideline-based care (43.3% at baseline vs 86.7% at 6 months) and an inverse correlation between baseline disease control and symptom improvements (r=−0.82; P<.001) reported in a small cohort trial, suggesting that patients with the poorest asthma control may benefit more from the intervention [37].

Assessment of medication and health care service use in studies with moderate bias risks showed some evidence of improvements in medication adherence rates (45.6% at baseline to 85.3% postintervention) [37], greater medication requests, and more preventive health care visits [36]. Reductions in hospital visits were also noted in a majority of provider reports (>60%) in Acharya et al [29], which was rated as high risk of bias. In another RCT with risk of bias, hospitalization rate reduction at 12-month follow-up was greater among patients who received training in portal use alongside home visits than among patients with no home visit support [32].

COPD

In the 2 COPD studies, reductions in the monthly hospitalization trend, emergency visits, and outpatient appointments postintervention were reported in a large quasi-experimental trial with low risk of bias (slope change −4.38,−5.08, and −0.004 per 10,000 patients per month, respectively), but there were no changes in hospital readmissions [38]. A greater mean change in annual hospital admission days (1.21 vs 0.62) and occupied bed days (8.12 vs 3.38) was reported in intervention versus control in another low-bias risk observational study, suggesting more effective disease management [41]. However, there were no survival benefits as the difference in the 12-month mortality rates was nonsignificant (16.9% intervention vs 24.1% control; P=.22) [38]. The findings were based on only 2 observational studies and they should be interpreted cautiously.

2. Patient-Centeredness

Overall, evidence from the 12 studies on asthma and COPD indicates high levels of patient satisfaction. Although some short-term QoL improvements among patient portal users with asthma were noted, sustained engagement was limited. Although QoL improvements remained relatively stable in COPD, this evidence was drawn from only 1 observational study. GRADE assessments showed that the certainty of evidence was very low for both asthma and COPD.

Asthma

Out of the 11 papers reporting on patient-centeredness in asthma, 7 papers showed positive effects such as high patient satisfaction [29-31,34,36,37,39], but issues related to adherence and engagement were also identified. These studies ranged from high to moderate risk of bias. Satisfaction rates were as high as 93.9% in a pilot study by Mammen et al. (2021) [36]. While some studies showed that patients rated patient portals better than usual care or paper diaries [34,37], sustained use was a notable issue with a drop in usage within the first month [34]. Similar issues with long-term engagement were also reported in a medium risk-of-bias rated observational study by Morita et al [39], with just over half of the participants (57.7%) using the tool at week 45, although the majority (65.2%) trusted the intervention. Alladina et al [31] also reported that despite 66% of patients recommending the application and nearly half (49%) rating it as excellent, overall usage rates remained low. In Apter et al [33], most patients (62%) expressed low confidence that patient portals would improve provider communication. Also, Zjip et al [44] showed that perceived self-management, as measured by the Perceived Control of Asthma Questionnaire, remained unchanged postintervention (21.4-21.6; d=0.06; P=.84), suggesting that the platform had a limited impact on patient self-management perceptions.

Five papers showed modest but variable QoL improvements [30,33,36,37,44]. A small study by Zjip et al [44] used the EuroQol questionnaire incorporating a visual analogue scale and found a nonsignificant QoL increase from 53.5 (SD 17.5) to 67.4 (SD 18.9) points (d=0.56; P=.06). The Mini Asthma-Related Quality of Life Questionnaire was used by all other studies. Although Ahmed et al [30] reported short-term QoL improvements at 3 months in the intervention group (mean Mini Asthma-Related Quality of Life Questionnaire change=0.67, 95% CI 0.36-0.98), this study had high risk of bias and they found no changes at 6 and 9 months. All other studies were rated as moderate risk of bias, except Apter et al [33], which had a low risk of bias, and they found no long-term QoL change at 12 months [33]. Also, Mammen et al [37] found that asthma-related QoL improvements were related to asthma control only (r=0.80; P<.001) but not with lung function measured using $FEV1\%_{pred}$ (r=0.087; P=.648), suggesting that QoL changes were linked to subjective asthma control but not objective lung function measures.

XSL•FO
RenderX

COPD

In COPD, the observational study by Taylor et al [41] showed that health-related QoL was stable over time, and symptom burden remained unchanged. Engagement was higher in patients experiencing more disease exacerbations, with an average of 4 self-managed exacerbations annually compared with the overall group average of two [41]. Evidence related to asthma or COPD showed that the presence of respiratory disease was not significantly associated with self-efficacy, and the outcomes for the 2 respiratory conditions were not presented separately [43].

3. Equity

Overall, evidence from the 8 studies on asthma and COPD found that patient portal use may be influenced by a range of sociodemographic and access-related factors, including technology availability, language, and digital literacy. Evidence of subgroup differences was limited and inconsistent, but access barriers appeared to affect different user groups. GRADE assessments showed that the certainty of evidence was low for asthma and very low for COPD.

Also, nearly half of the papers excluded participants without access to technology or the internet [31,34,37,39,40,42,44], and 2 papers explicitly focused on younger age groups below 40 years or 45 years [36,37]. Although some papers recruited purposeful samples of participants from areas of high deprivation [32,33], and reported on differences by language [35], exclusion of individuals without internet access or digital skills limits the generalizability of findings.

Asthma

Six out of 8 papers with medium to low risk of bias reported asthma-related subgroup differences in patient portal access and use, but the indicators used varied significantly [32,33,35,37,39,40]. One small cohort study showed greater lung function improvements ($FEV1\%_{predicted}$) among smokers, males, and participants with a high school education or less [37]. In terms of compliance, females and participants with a minimum bachelor's degree had higher questionnaire completion rates, showing better engagement in a feasibility trial, while age, clinical characteristics, and ethnicity were unrelated [40]. Spanish speakers, particularly those with lower literacy, were found less likely to use patient portals than English speakers in a study with low risk of bias (56% vs 42%; low-literacy groups=69% vs 49%). However, nearly a quarter in both groups reported that they would use the internet more if they had more knowledge or had access (English vs Spanish speakers=25% vs 26% and 25% vs 23%, respectively), highlighting access barrier issues among all participants [35]. In contrast, Apter et al [32] found no significant differential benefit from the patient portal with home visit across any patient subgroups, including participants with different primary languages (Spanish vs English) [32]. In an observational study, limited access and personal choice were reported as factors influencing technology engagement, as patients with chronic illness or prior hospitalizations were found less likely to use the portal [33]. A larger observational study with medium risk of bias by Morita et al [39] showed that older age (50 years and older) was associated with higher portal use, although only half of all participants reported smartphone ownership and use (55.2%) [39].

COPD

One COPD paper by Taylor et al [41] showed no subgroup differences in patient portal use between participants from socioeconomically deprived and nondeprived areas [41]. Tossaint-Schoenmakers et al [43] showed lower portal usability in those with asthma or COPD than in those with no chronic disease (b=−3.630; P=.02) and among those with higher education and age, suggesting the need for adaptive implementation strategies.

4. Efficiency

Overall, evidence from 5 studies on asthma suggested that patient portals may offer perceived improvements in aspects of care efficiency, such as enhanced communication and streamlined clinical interactions. However, these findings were based on subjective measures reported by small-scale studies, with limited objective evaluation of resource usage or cost-effectiveness. GRADE assessments showed that the certainty of evidence was very low for asthma.

Three out of the 5 papers showed positive impacts on asthma care [31,34,44], and communication, where 44% of participants in Alladina et al [31] reported improved disease-related discussions with their care providers, and 37% agreed that it helped their appointment run more smoothly. In terms of task completion, the feasibility of internet diary entry was rated similar to paper diaries, while lung function monitoring was faster (median 2 minutes) [34]. However, the overall evidence was considered weak due to high risk of bias.

Economic evaluation using descriptive metrics showed a mean cost of the patient portal of US $186.52 per participant over 6 months and an average nursing time of approximately 45 minutes per visit [36,37]. However, these findings were from studies conducted in the United States only and objective cost-effectiveness and resource usage outcomes were rarely assessed.

5. Safety

Overall, 5 papers included safety reports and all were on asthma. Evidence indicated that patient portals were generally perceived as safe, with no reports of significant adverse events. However, the absence of robust and standardized safety assessments limits confidence in these findings. GRADE assessments showed that the certainty of evidence was very low for asthma.

Four out of the 5 studies with high to moderate risk of bias reported positive effects with no safety-related concerns noted by the participants or by the group involved in technology implementation [29,34,36,44]. There were also no adverse events reported in a small experimental study by Zijp et al [44]. Privacy concerns were reported by 16% of participants in Apter et al [33], and Cruz-Correia et al [34] noted that internet connectivity issues were most frequently reported.

XSL•FO
RenderX

6. Timeliness

Overall, evidence from the 5 studies on asthma and COPD found that patient portals may support more timely care processes, including faster symptom reporting and improved communication in asthma. However, these findings were based on a small number of studies with varying designs and outcomes. In particular, only 1 observational study reported on COPD, and it noted longer time to hospitalization. GRADE assessments showed that the certainty of evidence was very low for both asthma and COPD.

Asthma

Three papers reported on timeliness in asthma and they demonstrated some positive effects [34,36,37]. Functions such as remote symptom logs–facilitated timely care were reported in a small pilot study, and patients received 100% response from care providers to follow-up requests [36]. In an RCT with high risk of bias, Cruz-Correia et al [34] showed that patient portals reduced symptom-reporting delays and eliminated bulk entry errors attributed to paper-based tools.

COPD

A large observational study rated as low risk of bias by Taylor et al [41] reported on time to adverse events in COPD and found that the patient portal group went longer without COPD or respiratory-related hospitalization or death than controls without patient portal access (335 days vs 155 days), with a 26% lower risk in the intervention group (hazard ratio 0.74; P=.047).

Discussion

Interpretation and Main Findings

This systematic review found that the current evidence base in asthma and COPD remains methodologically weak and insufficient. Evidence on effectiveness showed some increases in disease control, medication adherence, and preventive service use, but limited evidence on long-term health benefits and survival was noted. Findings related to patient-centeredness revealed high satisfaction and some quality-of-life gains, but this did not translate to sustained engagement or clinical benefits. We also noted exclusion of participants experiencing digital health disparities and access barriers. When assessed collectively, the body of evidence across the IOM domains showed an overall high risk of bias, imprecision, and inconsistencies, resulting in low to very low certainty. Evidence on COPD is particularly limited.

This review adds to the existing literature by synthesizing available studies on the specific impacts of patient portals on quality of care for asthma and COPD. Although clear positive effects (eg, improvements in care through patient portal use) were not identified, these findings remain important in highlighting the current lack of strong evidence and identifying priorities for future research and quality-of-care improvements. This review cautions against a reliance on patient portals for clinical management of asthma and COPD and highlights that more methodologically robust primary studies are needed to identify which portal functions may yield clinically meaningful benefits for patients with respiratory disease.

Comparison With Existing Literature

We captured a modest number of papers on patient portal use in asthma and COPD (n=16), which highlights an emerging body of evidence, particularly for asthma, which was reported in the majority of the studies included. However, comprehensive evaluation of their impacts across defined quality domains for both conditions remains sparse, reflecting broader concerns that the implementation of health technologies may be outpacing the evidence supporting their benefits [53]. In asthma, evidence of positive disease control is in line with a previous review of wider digital health interventions, which found benefits in symptom management, medication adherence, and reduced emergency visits [18]. Although we also noted some benefits in markers of effectiveness, such as improved provider care and greater symptom reductions in those with the poorest baseline control, these advantages coincided with cosupport strategies such as home visits and portal use training. Facilitation strategies such as structured onboarding, technical support, along with provider encouragement are established best practices in patient portal implementation and drivers of sustained engagement [24,54], and our findings indicate that they may remain important. However, this evidence is primarily driven by methodologically weak studies and must be interpreted with caution.

The evidence on patient-centeredness also varied, but patient portal satisfaction was generally high. However, this did not translate to sustained use as the attrition rate remained high in both short-term (ie, 1 month) and long-term follow-up. These patterns suggest an initial positive patient response to patient portals but with challenges realted to continued engagement, a trend also reported in the wider digital health literature [55,56]. This could be due to reasons such as lack of motivation or usability issues, which need to be further investigated [56]. In asthma, QoL improvements were limited to the short term and not sustained over longer periods, but it remained unchanged in COPD. However, the limited evidence on COPD restricts the generalizability of findings and the strength of evidence.

In terms of equity, we found a notable exclusion of individuals with low technology literacy and access, reinforcing the evidence of underrepresentation of certain disadvantaged populations in digital health studies, which is a known concern [57]. Exclusion of certain user groups may have biased the study findings and led to an overestimation of the effectiveness of these interventions. Although some of the studies we reviewed recruited a purposeful sample of marginalized groups and reported some access inequalities and usage barriers, these issues appear to have affected both disadvantaged and nondisadvantaged populations equally [35]. Therefore, disparities in digital health may persist across diverse populations, requiring concerted efforts to improve accessibility and engagement [58]. Notably, 1 paper showed lower patient portal usability among individuals with respiratory disease and with higher education [43]. Although people with a higher disease burden are more likely to use technology to manage their health [59], our findings suggest that clinical need alone does not drive uptake, and education does not guarantee digital competency. This complexity is acknowledged in the literature, which highlights how multiple factors such as prior technology

XSL•FO
RenderX

exposure, personal motivation, and support availability can interact to affect digital access, literacy, and skills needed to effectively engage with health care technologies [60,61]. Therefore, the need to involve participants from different backgrounds at every stage from research through to delivery and to coproduce interventions with them may be important to ensure that technological solutions are equitable and truly responsive to the needs of different user groups [62].

Regarding safety, limited evidence showed that patient portals were generally considered safe, with some reports of privacy concerns in asthma. Previous reviews highlight that security and privacy are key barriers restricting successful uptake of interventions providing EHR access, but we could not find strong evidence to support this [18,63,64]. However, this evidence was derived from studies with low to moderate methodological quality and includes small-scale, pilot trials, so the observed lack of association may be due to study limitations instead of an actual absence of safety concerns.

Strengths and Limitations

This study systematically synthesizes diverse evidence from studies using varied quantitative methods, endpoints, and geographical contexts using the IOM framework for a structured comparison. To the best of our knowledge, this is among the first studies to provide a comprehensive overview of the current state of evidence on patient portals in asthma and COPD by focusing on multiple health care quality domains. We conducted a narrative synthesis in line with SwiM guidance, grouping outcomes according to the IOM quality-of-care domains. This required subjective judgment, particularly where studies reported overlapping or nonstandardized outcomes, which need to be considered when interpreting results. We combined PRISMA and SWiM methods alongside the GRADE approach to assessing evidence certainty. All of these processes require interpretation of data and findings, and these may differ across groups. Nonetheless, we provide access to the relevant documentation in the appendices to aid transparency.

There are also other important limitations that warrant careful consideration. First, the current evidence is driven by small, often underpowered feasibility studies. Particularly for COPD, there is a lack of substantial evidence and high-quality studies. Although we conducted a comprehensive literature search using a broad set of key terms, it is possible that we may have missed some EHR-linked patient portal interventions used in asthma and COPD, which are not specifically indexed as such, or because they were not written in English. Also, all included studies were published post-2007, likely reflecting the broader adoption of patient portals from 2006, initially in primary care and outpatient settings [24]. As the use of digital health technologies appears to be increasing, especially for asthma and COPD, more evidence would be beneficial [65]. Second, some of the studies relate to the same trial, which further restricts the number of primary studies available and limits the generalizability of our findings. Third, due to considerable heterogeneity across the interventions and outcome measures reported, it is difficult to identify which components of the patient portals, if any, influence the observed effects. Variations in portal design and functionalities also limit generalizability of the findings, and sustained engagement was a noted issue. Finally, outcomes related to safety and timeliness were underexplored and mostly subjectively reported. The lack of evidence on these important care domains further restricts the ability to fully determine whether patient portals meaningfully contribute to respiratory disease care.

Implications for Research, Practice, and Policy

The findings of this review highlight the need for a stronger evidence base to support the implementation of patient portals in respiratory disease pathways. There is growing interest in the use of patient-facing digital health interventions, as evidenced by the recent National Institute for Health and Care Excellence guideline recommending several tools for COPD self-management (including 1 intervention included in this review) [41,66], and similar guidance for asthma is also underway [67]. However, the available research does not yet offer definite support for their routine use, highlighting the need for continued rigorous evaluation to support implementation efforts. In particular, the need for real-world evidence addressing digital engagement barriers is important. The recent NHS long-term plan may offer opportunities to integrate respiratory care–specific functions within national digital health platforms such as the NHS App [68], which can mitigate some of the access and engagement barriers we noted. For example, patients in England can already view test results, access care advice (including those specific to respiratory symptoms), and manage their medication list through the NHS App [68]. Therefore, facilitation efforts enabling wider use of national digital portals may offer some of the respiratory disease–specific benefits and allow coordinated management of multiple health conditions. However, as a precursor to this, further methodologically robust studies and adequately powered RCTs are crucial to address the limitations of the existing evidence base and to establish the validity of the reported benefits. This could include studies mapping specific functions to clinical outcomes and those with a longer follow-up period, including diverse groups such as adults and pediatric populations with asthma. Data collection approaches also need to be intersectional, recruiting a wide range of patient characteristics and across the various quality domains to identify user groups that would most benefit from these services.

Conclusions

This comprehensive assessment of primary studies does not provide support for meaningful improvements in the quality of care for asthma and COPD using patient portals. The evidence on COPD is restricted to a few large observational trials. While studies report some disease-specific gains (eg, modest asthma control improvements and reduced hospitalizations in COPD), these findings are based on a small number of predominantly low-quality studies, which limits confidence in the overall evidence. This review strengthens the existing literature by offering a synthesis of the most up-to-date findings on the impacts of patient portals in asthma and COPD, addressing the fragmented focus of prior reviews that grouped various technologies together. However, the available research provides insufficient confirmation of meaningful care quality improvements, suggesting that reliance on patient portals for

XSL•FO
RenderX

routine clinical management of asthma and COPD should be approached cautiously. If patient portals are to play a greater role in supporting self-management and quality of care, further studies on impacts and equity are required.

Acknowledgments

No AI tools were used in the preparation of this submission.

Funding

This research forms part of a doctoral thesis and is supported by the President's PhD scholarship from Imperial College London. ALN is funded by the National Institute for Health and Care Research (NIHR) North West London Patient Safety Research Collaboration (NIHR NWL PSRC), with infrastructure support from the Imperial NIHR Biomedical Research Centre. AM is supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Northwest London. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. No artificial intelligence tools were used in the preparation of this submission.

Data Availability

No new data were generated or analyzed in this systematic review study. All data supporting the findings are derived from previously published studies, which are cited in the paper.

Conflicts of Interest

None declared

Multimedia Appendix 1
Full search strategy and additional study details.
[PDF File, 219 KB - ojphi_v18i1e83483_app1.pdf]

Multimedia Appendix 2
Key characteristics of the included studies evaluating patient portals in adult asthma and chronic obstructive pulmonary disease (COPD). The table details the key characteristics of the 16 included papers, which include the study design, population included, intervention features, and comparator.
[PDF File, 201 KB - ojphi_v18i1e83483_app2.pdf]

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 33 KB - ojphi_v18i1e83483_app3.docx]

Checklist 2
PRISMA abstract checklist.
[DOCX File, 270 KB - ojphi_v18i1e83483_app4.docx]

Checklist 3
SWiM checklist.
[DOCX File, 20 KB - ojphi_v18i1e83483_app5.docx]

References

1. Momtazmanesh S, Moghaddam SS, Ghamari SH, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. EClinicalMedicine 2023 May;59:101936. [doi: 10.1016/j.eclinm.2023.101936]
2. RightCare asthma. NHS England. URL: https://www.england.nhs.uk/long-read/rightcare-asthma-scenario/ [accessed 2025-08-08]
3. Asthma. British Thoracic Society. URL: https://www.brit-thoracic.org.uk/quality-improvement/clinical-resources/asthma/ [accessed 2025-08-08]
4. Quality standards and indicators. : National Institute for Health and Care Excellence (NICE); 2015 URL: https://www.nice.org.uk/guidance/qs10/documents/briefing-paper [accessed 2025-08-08]
5. Respiratory disease. NHS England. URL: https://www.england.nhs.uk/ourwork/clinical-policy/respiratory-disease/ [accessed 2025-08-10]

XSL•FO
RenderX

6. Catching our breath: time for change in respiratory care. : Royal College of Physicians; 2025 URL: https://www.rcp.ac.uk/media/5e5mkucn/cqid_nrap_state-of-the-nation_catching-our-breath_report-final.pdf [accessed 2025-08-10]
7. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020 Nov 11;22(11):e23851. [doi: 10.2196/23851] [Medline: 33174851]
8. Global strategy for asthma management and prevention: 2024 update. : Global Initiative for Asthma; 2024 URL: https://ginasthma.org/wp-content/uploads/2024/05/GINA-2024-Strategy-Report-24_05_22_WMS.pdf [accessed 2026-2-10]
9. Global strategy for the diagnosis, management, and prevention of COPD (2026). : Global Initiative for Chronic Obstructive Lung Disease; 2026 URL: https://goldcopd.org/2026-gold-report-and-pocket-guide/ [accessed 2026-02-10]
10. Reed ME, Millman A, Fireman B, et al. Use of Patient Portals by People With Long-Term Health Problems: Patient-Centered Outcomes Research Institute; 2019. URL: http://www.ncbi.nlm.nih.gov/books/NBK603846 [accessed 2026-02-10]
11. Fiks AG, Mayne SL, Karavite DJ, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973. [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
12. Banguilan KL, Sonnenberg F, Chen C. Physicians' perspectives on inpatient portals: systematic review. Interact J Med Res 2022 Nov 15;11(2):e39542. [doi: 10.2196/39542] [Medline: 36378521]
13. Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM. Patient-centered digital health records and their effects on health outcomes: systematic review. J Med Internet Res 2022 Dec 22;24(12):e43086. [doi: 10.2196/43086] [Medline: 36548034]
14. Norouzi Aval R, Rafatpanah H, Sarbaz M, Mousavi Baigi SF, Kimiafar K. Identification and classification features of patient portal, a systematic review. Health Sci Rep 2025 Mar 23;8(3):e70520. [doi: 10.1002/hsr2.70520] [Medline: 40129513]
15. Crossing the Quality Chasm: A New Health System for the 21st Century: National Academies Press; 2001. [doi: 10.17226/10027]
16. Neves AL, Smalley KR, Freise L, Harrison P, Darzi A, Mayer EK. Determinants of use of the care information exchange portal: cross-sectional study. J Med Internet Res 2021 Nov 11;23(11):e23481. [doi: 10.2196/23481] [Medline: 34762063]
17. Campbell K, Greenfield G, Li E, et al. The impact of virtual consultations on the quality of primary care: systematic review. J Med Internet Res 2023 Aug 30;25:e48920. [doi: 10.2196/48920] [Medline: 37647117]
18. Carini E, Villani L, Pezzullo AM, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021 Sep 8;23(9):e26189. [doi: 10.2196/26189] [Medline: 34494966]
19. Kinley E, Skene I, Steed E, Pinnock H, McClatchey K. Delivery of supported self-management in remote asthma reviews: a systematic rapid realist review. Health Expect 2022 Aug;25(4):1200-1214. [doi: 10.1111/hex.13441] [Medline: 35411670]
20. Chan A, De Simoni A, Wileman V, et al. Digital interventions to improve adherence to maintenance medication in asthma. Cochrane Database Syst Rev 2022 Jun 13;6(6):CD013030. [doi: 10.1002/14651858.CD013030.pub2] [Medline: 35691614]
21. Shaw G, Whelan ME, Armitage LC, Roberts N, Farmer AJ. Are COPD self-management mobile applications effective? A systematic review and meta-analysis. NPJ Prim Care Respir Med 2020 Apr 1;30(1):11. [doi: 10.1038/s41533-020-0167-1] [Medline: 32238810]
22. British Thoracic Society, National Institute for Health and Care Excellence. BTS/NICE/SIGN joint guideline on asthma: diagnosis, monitoring and chronic asthma management. Thorax 2025 Jun;80(7):416-424. [doi: 10.1136/thorax-2024-222910] [Medline: 40274410]
23. 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. [doi: 10.2196/jmir.2238] [Medline: 23183044]
24. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148. [doi: 10.2196/jmir.4255] [Medline: 26104044]
25. Alturkistani A, Qavi A, Anyanwu PE, Greenfield G, Greaves F, Costelloe C. Patient portal functionalities and patient outcomes among patients with diabetes: systematic review. J Med Internet Res 2020 Sep 22;22(9):e18976. [doi: 10.2196/18976] [Medline: 32960182]
26. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890. [doi: 10.1136/bmj.l6890] [Medline: 31948937]
27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906. [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
28. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
29. Acharya S, Sarraf R. A novel preventative solution for effective asthma management: a practical evaluation. Netw Model Anal Health Inform Bioinform 2017 Dec;6(1):15. [doi: 10.1007/s13721-017-0156-7]
30. Ahmed S, Ernst P, Bartlett SJ, et al. The effectiveness of web-based asthma self-management system, My Asthma Portal (MAP): a pilot randomized controlled trial. J Med Internet Res 2016 Dec 1;18(12):e313. [doi: 10.2196/jmir.5866] [Medline: 27908846]
31. Alladina J, Moschovis PP, Gandhi HN, et al. Observational study of the Amaze™ asthma disease management platform. Digit Health 2024 Oct 22;10:20552076241282380. [doi: 10.1177/20552076241282380] [Medline: 39479373]

XSL•FO
RenderX

32. Apter AJ, Localio AR, Morales KH, et al. Home visits for uncontrolled asthma among low-income adults with patient portal access. J Allergy Clin Immunol 2019 Sep;144(3):846-853. [doi: 10.1016/j.jaci.2019.05.030] [Medline: 31181221]
33. Apter AJ, Bryant-Stephens T, Perez L, et al. Patient portal usage and outcomes among adult patients with uncontrolled asthma. J Allergy Clin Immunol Pract 2020 Mar;8(3):965-970. [doi: 10.1016/j.jaip.2019.09.034] [Medline: 31622684]
34. Cruz-Correia R, Fonseca J, Lima L, et al. Web-based or paper-based self-management tools for asthma--patients' opinions and quality of data in a randomized crossover study. Stud Health Technol Inform 2007;127:178-189. [Medline: 17901611]
35. Localio AM, Klusaritz H, Morales KH, Ruggieri DG, Han X, Apter AJ. Primary language and the electronic health record patient portal: Barriers to use among Spanish-speaking adults with asthma. J Asthma 2022 Oct;59(10):2081-2090. [doi: 10.1080/02770903.2021.1989462] [Medline: 34634975]
36. Mammen JR, Java JJ, Halterman J, et al. Development and preliminary results of an electronic medical record (EMR)-integrated smartphone telemedicine program to deliver asthma care remotely. J Telemed Telecare 2021 May;27(4):217-230. [doi: 10.1177/1357633X19870025] [Medline: 31438761]
37. Mammen JR, Schoonmaker JD, Java J, et al. Going mobile with primary care: smartphone-telemedicine for asthma management in young urban adults (TEAMS). J Asthma 2022 Jan;59(1):132-144. [doi: 10.1080/02770903.2020.1830413] [Medline: 33064038]
38. Martínez Nicolás I, Lê Cook B, Flores M, et al. The impact of a comprehensive electronic patient portal on the health service use: an interrupted time-series analysis. Eur J Public Health 2019 Jun 1;29(3):413-418. [doi: 10.1093/eurpub/cky257] [Medline: 30544169]
39. Morita PP, Yeung MS, Ferrone M, et al. A patient-centered mobile health system that supports asthma self-management (breathe): design, development, and utilization. JMIR Mhealth Uhealth 2019 Jan 28;7(1):e10956. [doi: 10.2196/10956] [Medline: 30688654]
40. Rudin RS, Fanta CH, Qureshi N, et al. A clinically integrated mHealth app and practice model for collecting patient-reported outcomes between visits for asthma patients: implementation and feasibility. Appl Clin Inform 2019 Oct;10(5):783-793. [doi: 10.1055/s-0039-1697597] [Medline: 31618782]
41. Taylor A, Cushing A, Dow M, et al. Long-term usage and improved clinical outcomes with adoption of a COPD digital support service: key findings from the RECEIVER trial. Int J Chron Obstruct Pulmon Dis 2023 Jun 22;18:1301-1318. [doi: 10.2147/COPD.S409116] [Medline: 37378275]
42. To T, Lougheed MD, McGihon R, Zhu J, Gupta S, Licskai C. Does an mHealth system reduce health service use for asthma? ERJ Open Res 2020 Jul;6(3):00340-2019. [doi: 10.1183/23120541.00340-2019] [Medline: 32963990]
43. Tossaint-Schoenmakers R, Kasteleyn M, Goedhart A, Versluis A, Talboom-Kamp E. The impact of patient characteristics on their attitudes toward an online patient portal for communicating laboratory test results: real-world study. JMIR Form Res 2021 Dec 17;5(12):e25498. [doi: 10.2196/25498] [Medline: 34927593]
44. Zijp A, Versluis A, Joosse L, et al. Exploring the feasibility, acceptability, usability and safety of a digitally supported self-management intervention for uncontrolled asthma: a pre-post pilot study in secondary care. Digit Health 2024 Nov 5;10:20552076241292391. [doi: 10.1177/20552076241292391] [Medline: 39507014]
45. Neves AL, Freise L, Laranjo L, Carter AW, Darzi A, Mayer E. Impact of providing patients access to electronic health records on quality and safety of care: a systematic review and meta-analysis. BMJ Qual Saf 2020 Dec;29(12):1019-1032. [doi: 10.1136/bmjqs-2019-010581] [Medline: 32532814]
46. Goldzweig CL, Orshansky G, Paige NM, 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]
47. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis: John Wiley & Sons, Ltd; 2009.
48. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. The Ottawa Hospital Research Institute. URL: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [accessed 2025-8-10]
49. Alshabanat A, Zafari Z, Albanyan O, Dairi M, FitzGerald JM. Asthma and COPD Overlap Syndrome (ACOS): a systematic review and meta analysis. PLoS One 2015;10(9):e0136065. [doi: 10.1371/journal.pone.0136065] [Medline: 26336076]
50. RoB 2: a revised Cochrane risk-of-bias tool for randomized trials. Cochrane Methods Bias. URL: https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials [accessed 2025-08-08]
51. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
52. Murad MH, Mustafa RA, Schünemann HJ, Sultan S, Santesso N. Rating the certainty in evidence in the absence of a single estimate of effect. Evid Based Med 2017 Jun;22(3):85-87. [doi: 10.1136/ebmed-2017-110668] [Medline: 28320705]
53. Houston TK, Richardson LM, Cotten SR. Patient-directed digital health technologies: is implementation outpacing evidence? Med Care 2019 Feb;57(2):95-97. [doi: 10.1097/MLR.0000000000001068] [Medline: 30601341]
54. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
55. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

XSL•FO
RenderX

56. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
57. Seaborn K, Barbareschi G, Chandra S. Not only WEIRD but "Uncanny"? A systematic review of diversity in human-robot interaction research. Int J Soc Robot 2023 Mar 8:1-30. [doi: 10.1007/s12369-023-00968-4] [Medline: 37359427]
58. Craig S, McPeak KE, Madu C, Dalembert G. Health information technology and equity: applying history's lessons to tomorrow's innovations. Curr Probl Pediatr Adolesc Health Care 2022 Jan;52(1):101110. [doi: 10.1016/j.cppeds.2021.101110] [Medline: 34974991]
59. Yoon E, Hur S, Opsasnick L, et al. Disparities in patient portal use among adults with chronic conditions. JAMA Netw Open 2024 Feb 5;7(2):e240680. [doi: 10.1001/jamanetworkopen.2024.0680] [Medline: 38421645]
60. O'Connor A, Tai A, Carson-Chahhoud K. Isn't there an app for that? The role of smartphone and tablet applications for asthma education and self-management in adolescents. Children (Basel) 2021 Sep 9;8(9):786. [doi: 10.3390/children8090786] [Medline: 34572218]
61. 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]
62. Voorheis P, Petch J, Pham Q, Kuluski K. Maximizing the value of patient and public involvement in the digital health co-design process: a qualitative descriptive study with design leaders and patient-public partners. PLOS Digit Health 2023 Oct;2(10):e0000213. [doi: 10.1371/journal.pdig.0000213] [Medline: 37878566]
63. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021 Jun 2;46(2):192-204. [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]
64. Radparvar I, Ross MK. Barriers and facilitators of electronic patient portal uptake for asthma management. J Asthma 2023 Apr;60(4):635-646. [doi: 10.1080/02770903.2022.2087190] [Medline: 35726134]
65. Garcia Gonzalez-Moral S, Addis P, Sharma O, et al. Innovative technologies for asthma and COPD management in the community: scanning the horizon using rapid systematic review methods. BMJ Innov 2025 Aug:133-146. [doi: 10.1136/bmjinnov-2024-001330]
66. Digital technologies to support self-management of COPD: early value assessment. National Institute for Health and Care Excellence. 2024. URL: https://www.nice.org.uk/guidance/htg736/chapter/2-The-technologies [accessed 2026-02-18]
67. Digital platforms recommended to help people manage their asthma. National Institute for Health and Care Excellence. 2026. URL: https://www.nice.org.uk/news/articles/digital-platforms-recommended-to-help-people-manage-their-asthma [accessed 2025-02-18]
68. Fit for the future: 10 year health plan for England. NHS England. URL: https://www.england.nhs.uk/long-term-plan/ [accessed 2025-08-08]

Abbreviations

ACQ: Asthma Control Questionnaire
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
IOM: Institute of Medicine
NHS: National Health Service
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
PROSPERO: International Prospective Register of Systematic Reviews
QoL: quality of life
RCT: randomized controlled trial
SWiM: Synthesis Without Meta-analysis

XSL•FO
RenderX

Edited by E Mensah; submitted 03.Sep.2025; peer-reviewed by G Davlyatov, V Press; revised version received 19.Jun.2026; accepted 28.Jun.2026; published 24.Aug.2026.

Please cite as:
Kc S, Hopkinson NS, Neves AL, Majeed A, Laverty AA
Impact of Patient Portals on Asthma and Chronic Obstructive Pulmonary Disease–Related Quality-of-Care Outcomes: Systematic Review
Online J Public Health Inform 2026;18:e83483
URL: https://ojphi.jmir.org/2026/1/e83483
doi:10.2196/83483

© Sukriti Kc, Nicholas S Hopkinson, Ana Luisa Neves, Azeem Majeed, Anthony A Laverty. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 24.Aug.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing Patient Discomfort in Smartphone-Based Teledentistry From the Perspective of Dental Professionals: Qualitative Interview Study

Margaret Shenouda[1], BA; Boyen Huang[1], DDS, MHA, PhD; Emily C Schultz[1,2], RDH, MS; Mohamed Estai[3], MBBS, MSc, PhD; Sarbin Ranjitkar[4], BDS, BScDent (Hons), PhD; Jeff P Louie[5], MD; Patimaporn Pungchanchaikul[6], DDS, MClinDent, PhD

[1]School of Dentistry, University of Minnesota, 515 Delaware Street SE, Minneapolis, MN, United States
[2]College of Allied Health and Nursing, Minnesota State University, Mankato, MN, United States
[3]School of Human Sciences, University of Western Australia, Perth, Australia
[4]Adelaide Dental School, The University of Adelaide, Adelaide, Australia
[5]Department of Pediatrics, Medical School, University of Minnesota, Minneapolis, MN, United States
[6]Faculty of Dentistry, Khon Kaen University, Khon Kaen, Thailand

Corresponding Author:
Boyen Huang, DDS, MHA, PhD
School of Dentistry, University of Minnesota, 515 Delaware Street SE, Minneapolis, MN, United States

Abstract

Background: Mobile health (mHealth) represents a modality of teledentistry that has the potential to improve access to dental care. Given that patient reactions to dental procedures can influence both clinician experience and care delivery, assessing patient discomfort when smartphones are used to capture dental images for teledentistry examinations is crucial.

Objective: This study aimed to explore patient discomfort from the perspective of dental professionals using smartphone-based photography in teledentistry.

Methods: A qualitative study was conducted through group interviews with a sample (N=10) of dental professionals, all of whom had experience capturing dental photos using smartphones equipped with an mHealth app at dental clinics and research facilities in Thailand and the United States. Audio-recorded interviews were transcribed, coded through consensus, and analyzed thematically.

Results: The dental professionals, including dental specialists, general dentists, dental therapists, and dental students, reported minimal to no patient discomfort during smartphone-based dental photography. Key factors contributing to patient comfort during teledentistry encounters included clear communication, informed consent, and reassurances regarding privacy and data security.

Conclusions: The findings suggest that providing patients with clear information and managing expectations can help reduce discomfort in teledentistry encounters. Improving communication strategies may enhance patient comfort, support the adoption of mHealth practices, and optimize interactions between patients and health care providers. Future research directions are indicated, such as directly assessing patient discomfort and identifying strategies to further minimize discomfort in teledentistry. Additionally, expanding teledentistry training in dental education and professional development will better equip dental professionals to effectively use this technology, ultimately improving accessibility and patient-centered care in dentistry.

(Online J Public Health Inform 2026;18:e81163) doi:10.2196/81163

KEYWORDS

mobile health; mHealth; smartphone; teledentistry; discomfort; patient; dental professional

Introduction

Discomfort related to dental services can lead to dental care avoidance, impacting both oral and general health [1]. There are countless reasons why humans experience discomfort or distress, including camera shyness [2], a fear of being photographed [2], or concerns about their privacy or dignity being compromised [3]. Some people commonly experience anxiety or distress due to bleeding and/or pain caused by traumatic incidents [4]. Those emotional responses are different from dental anxiety, which is an anticipatory unpleasant feeling for dental procedures [1,5]. While evaluation scales varied among studies, meta-analyses estimated that up to 30% of children [6] and 15% of adults [7] experienced dental anxiety. Dental anxiety, fear, and phobia represent a spectrum of negative emotional reactions to dental environments and procedures [8],

XSL•FO
RenderX

with "discomfort" often used as an umbrella term to describe experiences that range from mild unease to excruciating distress [9]. In this study, patient discomfort encompasses all levels of unpleasant feelings related to dental care.

Diagnostic tools, such as radiographs [5,10], intraoral cameras [11], and smartphones [12], make image assessments possible for dental professionals. Previous research has reported that pediatric patients experienced more anxiety when their dental treatment involved radiographs [5]. In adults, elevated anxiety biomarkers were also observed among those undergoing radiographic examinations for the first time [10]. To enhance patient experience and diagnostic accuracy, minimizing patient discomfort during the diagnostic imaging procedure is critical. A potential approach to lower dental discomfort is using mobile health (mHealth) technologies [13].

The mHealth practice provides a way for health care professionals to incorporate mobile phones and other technological devices into their practices [14]. As a modality of teledentistry [15], mHealth can be used for teleconsultation, telediagnosis, telemonitoring, telesupport, and teleintervention [16]. Smartphone-based dental photography also helps with communication and documentation in clinical settings [17]. The implementation of mHealth increases accessibility to health care services for rural and underserved communities [18]. The practice of mHealth grew in popularity during the COVID-19 pandemic when social distancing measures were required [19]. Previous research has reported that dental professionals [20,21], patients [21], and caregivers [22] all consider mHealth to be feasible and useful for remote dental services.

Teledentistry is increasingly being integrated into everyday dental procedures [23]. As mHealth technology seldom requires inserting instruments, devices, or radiographic film into the patient's mouth, it is generally considered noninvasive. However, some patients have reported anxiety during video-based teledentistry consultations [24]. Earlier studies have demonstrated that most dental professionals could perceive patient discomfort in clinical settings [25] and also feel stressed managing patients with anxiety symptoms [26]. To the best of the authors' knowledge, no existing literature has specifically addressed patient and clinician discomfort associated with photography-based teledentistry using mHealth approaches. Nonetheless, implementing strategies to reduce patient discomfort remains crucial for ensuring effective and patient-centered dental care [1], including teledentistry [24].

Whether mHealth practice would or would not induce discomfort in dental patients and/or dental professionals remained unclear. Therefore, this study addressed the following research questions: (1) whether dental professionals perceived potential discomfort from their patients in response to their use of the mHealth practice and (2) how patient discomfort (if any) influenced the dental professionals' experience operating the smartphone and mHealth app to capture images of their patients' teeth. Dental professionals' perceptions were used as a proxy for patients' reactions, as they were present during patient interactions, and interviewing patients in the clinical and research settings was not feasible. Thus, this study aimed to investigate dental professionals' perspectives on patient discomfort when using a smartphone to capture images of patients' teeth for teledentistry examinations. This study also explored the influence of patient discomfort on dental professionals' photo-taking experience.

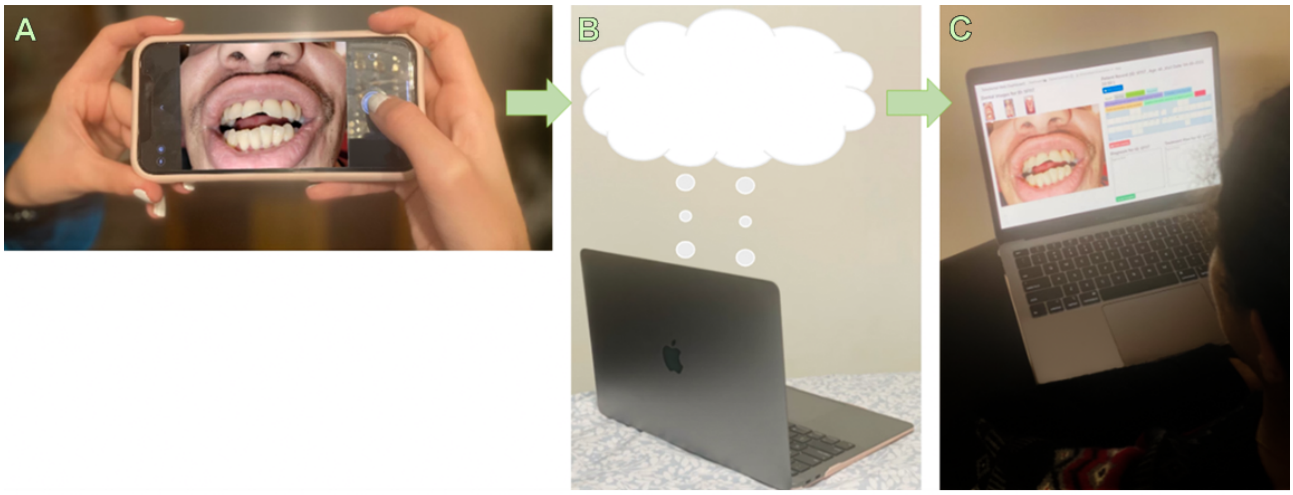
Methods

Overview

This qualitative study, forming a segment of a broader teledentistry research initiative, received approval from the University of Minnesota Institutional Review Board (STUDY00014736). The study was conducted across several sites in Minnesota, United States, and Khon Kaen, Thailand, including clinical facilities at the University of Minnesota, Khon Kaen University, and research venues at the Minnesota State Fair. In this paper, the term "patient" encompasses both dental patients recruited from clinics and fairgoers recruited at the State Fair research facilities, as both groups underwent the same teledentistry procedure. The quantitative and qualitative research methods used in the overarching project were recently reported [12,20]. In the quantitative segment, 11 trained and calibrated dental professionals used a Samsung smartphone equipped with an image acquisition app Teledental (CSIRO) to take dental photos (frontal, upper occlusal, and lower occlusal views) from consented dental patients and fairgoers and uploaded the images to a secure cloud storage for remote reviewers to assess [12]. Figure 1 illustrates the flow of the photographic acquisition and assessment in this mHealth model.

XSL•FO
RenderX

Figure 1. Illustration of the mobile health model: (A) a dental professional using the Teledental app to capture dental photos of a patient, (B) the data being uploaded from the smartphone to a secure cloud storage, and (C) a remote reviewer using the charting platform to review the dental photos asynchronously. The patient consented to have the images and data presented in the publication.



Ethical Considerations

This study was approved by the University of Minnesota Institutional Review Board (STUDY00014736). Written informed consent was obtained from all participants prior to data collection. All study personnel complied with institutional review board–approved protocols governing data access, confidentiality, and privacy. Participating dental professionals based in the United States received compensation (dental therapist: $40 USD; dental student: $15 USD); Thailand-based dental professionals did not, due to university restrictions on out-of-country payments.

Study Design

For the qualitative segment, we used a study design guided by the technology acceptance model (TAM) [27] and interpretivism [28] to examine the experiences of dental professionals and remote reviewers in using mHealth technology [20]. This work represents a portion of the qualitative segment and specifically explores patient discomfort from the perspective of dental professionals. Except for 1 dental professional (MS) who is an author of this paper, all 10 of the 11 dental professionals who captured dental images in the larger study were invited via email to participate in this qualitative study. All 10 professionals consented and attended a 60-minute interview conducted via Zoom (Zoom Communications). A combination of one-on-one and group interviews were conducted, with group size ranging from 1 to 6 interviewees, determined by their availability. Due to time zone differences, interviewees from the United States and Thailand were placed in separate groups. A total of 2 US-based researchers conducted the interviews in English, and all interviewees were proficient in English for the discussions. The study adhered to the Consolidated Criteria for Reporting Qualitative Research guidelines throughout its execution and reporting [29].

Research Team and Reflexivity

Each one-on-one or group interview was conducted by 2 interviewers: MS, a female honors college graduate with a bachelor of arts degree, and ECS, a female dental hygiene faculty member with a master of science degree. BH, a male senior dental academic and principal investigator with doctor of philosophy and doctor of dental surgery degrees, observed all sessions. Before the first interview, the interviewers received training to adopt a neutral interviewing approach, ensuring interviewees' perspectives were captured without bias. While most interviewees had no prior relationship with the interviewers, 3 were previously acquainted with an interviewer as dental professionals in a prior study led by the observer. The interviewers and observer introduced themselves first, and the objectives and details of this study were explained through the informed consent form and at the start of each interview.

Data Collection

The interviews were attended exclusively by the interviewees and researchers. During the Zoom sessions, the interviewers took field notes to record nonverbal cues and additional observations. All interviews were audio-recorded and securely stored on a cloud-based server. Following each interview session, interviewees were emailed the complete study instrument and given 2 weeks to provide additional feedback. Data saturation, defined as the point at which no new information emerged across 3 consecutive sessions [30], was a criterion applied to the broader qualitative project involving a larger participant pool [20]. Nevertheless, to ensure a comprehensive and diverse range of perspectives, all 10 dental professionals were interviewed across 4 sessions, irrespective of whether data saturation had been reached. No interviews were repeated for the same dental professionals.

The interview instrument consisted of guided questions designed to explore the dental professionals' perspectives. These questions were adapted from published resources [31,32] and rephrased to align with the context of this qualitative study and the specific mHealth technology used. A multidisciplinary panel of researchers, with expertise in qualitative study design, medical informatics, and dentistry, reviewed and finalized the interview questions to ensure their relevance and comprehensiveness. All dental professionals were asked 2 questions by the same interviewer. The first question was as follows: "How comfortable or uncomfortable did you feel about using the technology for dental trauma cases?" This question

XSL•FO
RenderX

focused on the dental professionals' own comfort levels when using the smartphone and mHealth app. The second question was as follows: "How might you have observed either potential distress or comfort from patients of the study?" The question was intended to capture the dental professionals' perspectives on their patients' comfort level. These 2 questions were part of a broader qualitative study [20] conducted, observed, and analyzed by the same research team. Although the larger study explored additional dimensions of mHealth use, only these 2 questions are reported here because they specifically addressed the anxious or emotional reactions experienced by dental professionals or patients during teledentistry procedures.

Data Analysis

The qualitative data from the interviews were analyzed using a consensus coding process to summarize responses and synthesize the results, with a deductive approach [33-36]. Prior to conducting the interviews, the coding team met via Zoom to calibrate their approach, ensuring consistency in interpretation across transcripts. Audio-recorded interviews were transcribed and independently reviewed by the coding team. Throughout the analysis, memo writing was used to record coder reflections, track emerging themes, and document the rationale behind coding decisions, thereby enhancing transparency. Intercoder reliability was addressed through a negotiated agreement strategy: coders had established procedures for resolving discrepancies in advance, and when disagreements occurred, they revisited the transcript, discussed interpretations, and reached consensus. The principal investigator (BH), who observed all interview sessions, was available to provide input when needed to ensure consistency. A hybrid coding framework was applied, combining deductive codes derived from the research questions and the TAM [27] with inductive codes that captured novel insights arising directly from the data. This flexible approach allowed the team to systematically examine anticipated themes while incorporating unexpected findings that enriched understanding of participants' experiences [33,37]. After all interviews were completed, the research team convened via Zoom to review the transcripts collectively, identify recurring themes, and extract supporting evidence, including descriptive terms and illustrative quotes from participants. The transcripts and findings were not shared with interviewees for review.

Results

Overview

A total of 4 interviews were conducted. All 10 dental professionals participated, 4 from the United States (coded U1-U4) and 6 from Thailand (coded T1-T6). Although U1 to U4 were initially intended to be interviewed together, scheduling constraints led to U1 and U4 being interviewed individually and U2 and U3 together, whereas T1 to T6 were all available simultaneously and, as affiliates of the same institution in Thailand, participated in a single group interview. Table 1 lists the interview sessions, professions, genders, and countries of the dental professionals.

The thematic analysis process, in which patterns or themes are identified within qualitative data revealed 2 main themes from the interviewees' responses: professional familiarity and confidence with mHealth technology, and patient reactions to smartphone imaging in dental care (Table 2).

Table . Interview sessions, professions, genders, and countries of the interviewees; 2 one-on-one interviews (sessions 1 and 3) and 2 group interviews (sessions 2 and 4) were conducted, with the group arrangement determined by the interviewees' time availability.

Interview session	Profession	Gender	Country	Interviewee code
1	Dental student	Male	United States	U1
2	Dental therapist	Female	United States	U2
2	Dental student	Male	United States	U3
3	Dental student	Female	United States	U4
4	General dentist	Female	Thailand	T1
4	General dentist	Female	Thailand	T2
4	General dentist	Female	Thailand	T3
4	General dentist	Male	Thailand	T4
4	General dentist	Male	Thailand	T5
4	Dental specialist	Male	Thailand	T6

XSL•FO
RenderX

Table . Main themes, definitions, subthemes (code words), and representative transcript excerpts that contributed to theme construction.

Main theme	Definition	Subtheme	Example from transcripts
Professional familiarity and confidence with mHealth technology	How dental professionals' prior experience and daily use of mobile phones shaped their comfort and perceived ease in adopting smartphone-based imaging	Confident adoption due to everyday use	• "Everyone seemed comfortable with...using the technology. We use mobile phone everyday all the time so we are comfortable with the technology."
		Personal ease contrasted with anticipated challenges for others	• "I think it was very easy for me. I think it was something that I felt very comfortable being able to use, taking the images, making sure the images came out well. I felt very comfortable doing that. I know it may be more challenging for people who are not as familiar with being able to use that technology."
Patient reactions to smartphone imaging in dental care	How dental professionals perceived patients' emotional responses to smartphone-based imaging	Patients appeared at ease once informed	• "Some patients were like 'oh yeah, this is pretty cool,' like it's the next step in doing things through electronics. I thought I got a lot of patients that felt at ease with this." • "I feel study participants [felt okay] once they understand and we explain to them the reason and what we're doing, and they're aware of it." • "We did not notice any anxiety or discomfort; it is like a normal part of an examination and there is no pain when taking photographs. • Because we asked the patients to sign the consent and they already gave consent prior to taking the photos, they did not notice any discomfort or anxiety from patients. Their regular dental [exam] is more invasive than taking the photos."
		Patients expressed unease linked to privacy, intimacy, or cultural context	• "I work with patients that are immigrants or potentially illegal immigrants and I think they get really nervous, even if you're saying you're not going to be in the photo at all, things like that." • "The mouth is a very intimate place for a lot of people. Letting people look in your mouth, especially if you have insecurities or are concerned about something or there's something that you know is a problem but you've been putting it off for a variety of reasons is difficult."

XSL•FO
RenderX

Professional Familiarity and Confidence With mHealth Technology

A total of 9 out of 10 dental professionals reported feeling comfortable using the mHealth app and taking dental photos. Several dental professionals attributed this to their familiarity with smartphones, as U4 and T4 noted they use mobile phones daily. U4 also expressed ease with both the app and photography process. Factors such as having a newer phone (stated by T4) and proper lighting and retraction (stated by U1) contributed to the feasibility of the process. However, U2 mentioned challenges with consistency, requiring a refresher before using the app due to infrequent use.

Despite overall comfort, 3 dental professionals highlighted specific concerns. U2, based in the United States, described using a cell phone in a dental setting as "invasive," particularly when working with privacy-conscious patients. This discomfort stemmed from introducing the phone and ensuring patients understood their data were secure, as stated by U2:

> *Whipping out a cell phone is kind of always uncomfortable. Especially with some of the patient types that I work with where they might be a little bit more concerned about privacy.*

U2 also noted a lack of prior experience with intraoral photography. T4 emphasized the importance of restricting data access to the research team, while T3, another dental professional based in Thailand, worried about appearing less competent in front of the patient when relying on an app for diagnosis.

Patient Reactions to Smartphone Imaging in Dental Care

The majority of dental professionals observed no noticeable discomfort among patients during photo-taking. T1 suggested that dentists can help patients become accustomed to photography as part of dental examinations by adopting the approach of "we take a photo and show it to them." U2 noted observing minimal discomfort, describing patients as "fine" but acknowledged some initial hesitancy, especially among immigrant patients who might harbor concerns about privacy or identification. U2, a US-based dental professional, noted that immigrant patients sometimes fear being photographed, even when assured they won't appear in the image (see Table 2).

Despite this, U2 mentioned that some patients viewed technology integration as a necessary advancement. U3, acting as a research recruiter at the Minnesota State Fair, observed that patients were generally comfortable by the time they arrived, perhaps because they volunteered or were not in a dental facility for dental care. However, recruiting participants posed a challenge, echoing common issues with dental appointment attendance. U4 found that providing more detailed information and setting clear expectations reduced patient discomfort, making them "less anxious" and "more willing" to participate.

Dental professionals also speculated why patient discomfort was minimal. T1 noted that patients in emergency trauma situations were generally uncooperative and preferred faster tools such as intraoral cameras. Conversely, patients with nonemergency issues were described as "very comfortable." T2 and T4 believed that obtaining informed consent and explaining the photo-taking process alleviated patient concerns. They emphasized that the procedure was painless and less invasive than standard dental examinations. T4 noted that patients seemed comfortable, particularly when dental photos were shared with them afterward to facilitate treatment discussions.

Clear communication was a recurring theme in reducing patient discomfort. T4 and U1 highlighted the importance of explaining the exclusion of identifiable features in photos. U1 stated the following:

> *Once we explained that [photos] wouldn't be of any identifiable features, most patients were comfortable.*

Transparency about the process helped foster trust and alleviate initial apprehension.

Discussion

Principal Findings

The overall consensus from dental professionals was that patients did not experience discomfort during smartphone-based dental photography. Although the telesupport function of some mHealth technologies could reduce patient discomfort [13], this study demonstrated that the asynchronous telediagnosis function of the mHealth technology could also minimize patient discomfort. Of further note, a prior study has reported a low level of dental patient anxiety from synchronous video consultation [24]. Thus, both synchronous and asynchronous teledentistry can decrease patient discomfort when dental professionals use these practices to communicate and examine patients.

One of the reasons for this finding is likely to include the physical safety characteristics of the mHealth model. Earlier studies have suggested that the use of X-rays in dental treatment increased patient anxiety, most likely due to the procedure's uncomfortable nature [5], apprehension toward radiation exposure [5], and the perceived lack of control during image acquisition [10]. Compared to X-rays, taking photos using a smartphone does not inflict any pain and is not invasive. Hence, patients become less uncomfortable when receiving a teledentistry procedure than undergoing a radiographic examination. Another possible reason patients did not express discomfort may have been the communication provided by dental professionals before the teledentistry procedure. Dental professionals informed patients about what the dental photos would include and how they would be used, which likely contributed to patient comfort. This aligns with the recommendation from literature, which emphasizes obtaining clear informed consent and ensuring confidentiality, particularly regarding who would have access to recorded materials and how they would be used, can enhance patient comfort [38] and support the effectiveness of psychotherapy [2]. Furthermore, dental professionals could show patients the images of their teeth after photography to help strengthen patient trust. These highlight the importance of health care provider–patient

XSL•FO
RenderX

communication and carry significant clinical implications for mHealth and teledentistry.

Beyond the clinical context, the findings of this study can also be interpreted through the lens of informatics and digital health system design. The TAM provided a useful framework for understanding how familiarity with smartphones influenced dental professionals' confidence and ease of adoption [27]. From a systems perspective, the mHealth model used here illustrates key principles of human-technology interaction: usability [39], transparency [40], and trust [40]. Prior informatics reviews have emphasized that clear communication of data flows and privacy safeguards is essential for user acceptance of digital health tools [40,41]. Integrating these insights into system design suggests that teledentistry platforms should prioritize intuitive interfaces [40], explicit consent protocols [41], and secure data management to enhance both health care provider and patient comfort [40,41]. In this way, the study contributes not only to clinical practice but also to broader discussions on how digital health systems can be designed to balance efficiency, accessibility, and patient-centered care.

In teledentistry procedures, smartphones and intraoral cameras are often used as alternatives to each other. A recent study found that patients reported greater comfort with smartphone-based teledentistry compared to intraoral cameras, although smartphone photography required more time to complete a similar procedure [42]. Using a smartphone for diagnostics may cause less discomfort than an intraoral camera because smartphones are familiar devices used daily by many people. In contrast, intraoral cameras are only found in clinical settings and can feel more intrusive. In cases of traumatic dental injuries, high-quality photographs during the initial visit are essential for long-term follow-up, which may be required for up to 5 years [43]. Patients will need to undergo photography during most follow-up visits. As mHealth apps on smartphones feel less intimidating than intraoral photography, they could help reduce patient discomfort and encourage regular checkups.

Only 1 interviewee raised a unique concern regarding the potential discomfort that smartphone-based photography might cause among immigrant patients. Although all patient photos were deidentified during the research process and patients were reassured about privacy and confidentiality protections, maintaining anonymity in photography-based research involving migrant minority groups remains an important consideration [44]. To address this, patient education programs could help reassure patients about the safety of teledentistry, while professional development for clinicians and administrators could reinforce best practices for protecting patient privacy. Additionally, dental photographs captured through mHealth approaches should be securely integrated into electronic health records to ensure confidential access and long-term storage. For patients who are uncomfortable with teledentistry, in-person appointments should remain an option to ensure continued access to dental care.

Dental professionals also found the mHealth teledentistry model to be feasible, reporting a high level of comfort with its use. This aligns with findings in medical education, where students with prior exposure to telemedicine tend to feel more confident using such technologies [45]. On the basis of these insights, we recommend integrating comprehensive teledentistry training into dental education and ongoing professional development. These programs should prioritize building core competencies in teledentistry, such as obtaining informed consent and maintaining patient confidentiality during virtual encounters. Incorporating teledentistry training into dental curricula and clinical protocols could enhance practitioner readiness and improve patient care in remote settings. Standardizing communication protocols and ethical practices, such as consent and confidentiality, can help establish trust and ensure quality care. As teledentistry becomes more widespread, these educational and policy shifts will be essential for ensuring its effective and responsible implementation in clinical practice.

The interpretation of this study's findings is subject to certain limitations. First, the small sample size of interviewees may restrict the ability to draw generalized conclusions, as a larger cohort might reveal additional perspectives. The pool of participants, consisting of 4 US-based and 6 Thailand-based dental professionals, may have introduced biases shaped by their unique clinical experiences, cultural backgrounds, patient populations, and photo-capturing environments. For instance, dental professionals assigned to the Minnesota State Fair research facilities were unlikely to encounter patients with emergency trauma cases, as such patients would typically seek care in clinics or hospitals rather than visiting the State Fair. Conversely, those working in clinical settings were more likely to manage patients presenting with pain, bleeding, or injuries and may have been more confident and experienced in clinical dentistry. Notably, privacy concerns among immigrant patients were raised by American dental professionals, whereas Thai dental professionals reported image concerns about appearing less competent in front of patients. These differences may reflect broader cultural expectations. An earlier study suggested that Thais were inclined to accommodate audience expectations, while Americans often balanced personal goals with audience interests [46]. Such cultural norms likely shaped how dental professionals perceived patient concerns and articulated their views. Given these contextual influences and the small, convenience-based sample, the transferability of findings is limited. Future research should purposely recruit larger and more diverse cohorts of dental professionals working with similar mHealth models to strengthen the depth, cultural breadth, and applicability of insights.

Another limitation of this study is its focus on patient discomfort as perceived by dental professionals, rather than directly assessing patient discomfort from the patients' perspectives. Given the small sample size and the varied clinical experiences of the interviewees, interpretations of patient discomfort may reflect biases stemming from dental professionals' subjective perceptions. While indirect assessments of patient discomfort from clinicians' viewpoints may not capture the full scope and intensity of patients' experiences, dentists are well positioned to observe and interpret discomfort [25], and they can also be affected by their patients' discomfort [26]. Therefore, investigating patient discomfort from dental professionals' perspectives remains a practical and informative approach, despite its limitations.

XSL•FO
RenderX

Future directions based on this study include exploring patient discomfort from the perspectives of both patients and caregivers, as well as identifying the factors contributing to discomfort associated with different teledentistry technologies. Additionally, developing and evaluating effective strategies to alleviate such discomfort is a critical area for future research. This study also underscores the importance of defining and cultivating core competencies in teledentistry, particularly in understanding and managing patient discomfort. Refining training curricula for current and future clinicians to incorporate these competencies will better prepare them to address patient concerns and maximize the effectiveness of teledentistry practices.

Conclusions

This study highlights the feasibility of smartphone-based photography for teledentistry, with dental professionals expressing comfort in using the mHealth app. Familiarity with smartphones and clear communication helped facilitate the process, although minor challenges such as maintaining consistency in app use and concerns about smartphone introduction in clinical settings were noted. Patients exhibited minimal discomfort, likely due to the noninvasive nature of smartphone photography and the role of informed consent in setting expectations.

Clinically, integrating smartphone-based photography into teledentistry requires clear communication strategies to enhance patient comfort and trust. Future research should directly assess patient discomfort during teledentistry encounters and explore strategies to further minimize discomfort. Expanding teledentistry training in dental education and professional development will better equip dental professionals to use this technology effectively, ultimately improving accessibility and patient-centered care in dentistry.

Acknowledgments

The authors acknowledge Minnesota Masonic Charities, the International Association for Dental Research, the University of Minnesota Center for Global Health and Social Responsibility, and the University of Minnesota School of Dentistry for their funding support. They also thank the individuals who participated in the interviews. In addition, the authors are grateful to Professor Katherine Klink for her input and support. Part of this work was presented in MS's honors thesis and at the University of Minnesota Undergraduate Research Symposium. Portions of this manuscript were rephrased with the assistance of ChatGPT (January 25 version, GPT-4-turbo), a generative artificial intelligence (AI) technology developed by OpenAI. The AI did not generate or edit any original written or visual content for this paper. All rephrased content was reviewed by the authors to ensure factual accuracy and the absence of plagiarism.

Funding

This study was supported by a Department of Pediatrics Driven to Discover Grant from Minnesota Masonic Charities, a Regional Development Program Grant from the International Association for Dental Research, a Global Health Seed Grant from the University of Minnesota Center for Global Health and Social Responsibility, and Faculty Start-up Funding for BH from the University of Minnesota School of Dentistry. The funding sources had no role in the design of the study and were not involved in its execution, data analysis, interpretation, or the decision to submit the results.

Data Availability

The datasets generated or analyzed during this study are not publicly available because participants did not provide written consent for their recordings to be shared, but the transcribed data are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BH (lead), ME, SR, JPL, PP
Data curation: ECS (lead), MS (supportive)
Funding acquisition: BH (lead), ME, SR, JPL, PP
Methodology: BH (lead), ECS, ME, SR, PP
Project administration: ECS (lead), MS (supportive)
Resources: BH (equal), JPL (equal), PP (equal)
Supervision: BH (equal), PP (equal)
Validation: BH (equal), ECS (equal)
Visualization: MS (lead), BH (supportive)
Writing – original draft: MS
Writing – review & editing: BH (lead), ECS, ME, SR, JPL, PP

Conflicts of Interest

None declared.

Checklist 1

XSL•FO
RenderX

COREQ checklist.
[PDF File, 447 KB - ojphi_v18i1e81163_app1.pdf]

References

1. Hoffmann B, Erwood K, Ncomanzi S, Fischer V, O'Brien D, Lee A. Management strategies for adult patients with dental anxiety in the dental clinic: a systematic review. Aust Dent J 2022 Mar;67 Suppl 1(Suppl 1):S3-S13. [doi: 10.1111/adj.12926] [Medline: 35735746]
2. Falzone RL, Hall S, Beresin EV. How and why for the camera-shy: using digital video in psychiatry. Child Adolesc Psychiatr Clin N Am 2005 Jul;14(3):603-612. [doi: 10.1016/j.chc.2005.02.006] [Medline: 15936676]
3. Macintosh T. Ethical considerations for clinical photography in the global South. Dev World Bioeth 2006 May;6(2):81-88. [doi: 10.1111/j.1471-8847.2006.00142.x] [Medline: 16594971]
4. Turner E, Robinson DM, Roaten K. Psychological issues. Phys Med Rehabil Clin N Am 2023 Nov;34(4):849-866. [doi: 10.1016/j.pmr.2023.05.005] [Medline: 37806702]
5. Mahiepala NA, Phan VL, Kieu KD, Koppen JP, Hussain BH, Huang B. Influencing factors of paediatric dental anxiety levels in an undergraduate dental clinic. Eur J Paediatr Dent 2015 Jun;16(2):159-162. [Medline: 26147825]
6. Sun IG, Chu CH, Lo EC, Duangthip D. Global prevalence of early childhood dental fear and anxiety: a systematic review and meta-analysis. J Dent 2024 Mar;142:104841. [doi: 10.1016/j.jdent.2024.104841] [Medline: 38246307]
7. Silveira ER, Cademartori MG, Schuch HS, Armfield JA, Demarco FF. Estimated prevalence of dental fear in adults: a systematic review and meta-analysis. J Dent 2021 May;108:103632. [doi: 10.1016/j.jdent.2021.103632] [Medline: 33711405]
8. Avramova NT. Dental fear, anxiety, and phobia; causes, diagnostic criteria and the medical and social impact. J Mind Med Sci 2022;9(2):202-208. [doi: 10.22543/2392-7674.1348]
9. Lindsay SJ, Wege P, Yates J. Expectations of sensations, discomfort and fear in dental treatment. Behav Res Ther 1984;22(2):99-108. [doi: 10.1016/0005-7967(84)90098-6]
10. Dantas Costa E, Cano Martins LA, Carvalho Chaves-Junior S, et al. Patients´ anxiety related to dental radiographic examinations: levels, perception and management. Int J Odontostomat 2021;15(4):971-979. [doi: 10.4067/S0718-381X2021000400971]
11. Vetchaporn S, Rangsri W, Ittichaicharoen J, Rungsiyakull P. Validity and reliability of intraoral camera with fluorescent aids for oral potentially malignant disorders screening in teledentistry. Int J Dent 2021;2021:6814027. [doi: 10.1155/2021/6814027] [Medline: 34745263]
12. Huang B, Estai M, Pungchanchaikul P, et al. Mobile health assessment of traumatic dental injuries using smartphone-acquired photographs: a multicenter diagnostic accuracy study. Telemed J E Health 2024 Oct;30(10):2592-2600. [doi: 10.1089/tmj.2024.0012] [Medline: 38946689]
13. Väyrynen E, Hakola S, Keski-Salmi A, Jämsä H, Vainionpää R, Karki S. The use of patient-oriented mobile phone apps in oral health: scoping review. JMIR Mhealth Uhealth 2023 Sep 6;11:e46143. [doi: 10.2196/46143] [Medline: 37672331]
14. Davies A, Mueller J. Developing Medical Apps and mHealth Interventions: A Guide for Researchers, Physicians and Informaticians: Springer; 2020.
15. ADA policy on teledentistry. American Dental Association. 2020. URL: https://www.ada.org/en/about/governance/current-policies/ada-policy-on-teledentistry [accessed 2026-03-23]
16. Mariño RJ, Uribe SE, Chen R, Schwendicke F, Giraudeau N, Scheerman JF. Terminology of e-oral health: consensus report of the IADR's e-Oral Health Network Terminology Task Force. BMC Oral Health 2024 Feb 28;24(1):280. [doi: 10.1186/s12903-024-03929-z] [Medline: 38419003]
17. Hardan LS, Moussa C. Mobile dental photography: a simple technique for documentation and communication. Quintessence Int 2020;51(6):510-518. [doi: 10.3290/j.qi.a44365] [Medline: 32424377]
18. Steinman L, Heang H, van Pelt M, et al. Facilitators and barriers to chronic disease self-management and mobile health interventions for people living with diabetes and hypertension in Cambodia: qualitative study. JMIR Mhealth Uhealth 2020 Apr 24;8(4):e13536. [doi: 10.2196/13536] [Medline: 32329737]
19. Harris B, Ajisola M, Alam RM, et al. Mobile consulting as an option for delivering healthcare services in low-resource settings in low- and middle-income countries: a mixed-methods study. Digit Health 2021 Aug 19;7:20552076211033425. [doi: 10.1177/20552076211033425] [Medline: 34777849]
20. Schultz EC, Huang B, Shenouda M, et al. Perspectives of frontline clinicians and end-line users on smartphone-based photography for assessing traumatic dental injuries: focus group interview study and thematic analysis. JMIR Hum Factors 2026 Jan 26;13:e82668 [FREE Full text] [doi: 10.2196/82668] [Medline: 41587769]
21. Tokede B, Yansane A, Ibarra-Noriega A, et al. Evaluating the impact of an mHealth platform for managing acute postoperative dental pain: randomized controlled trial. JMIR Mhealth Uhealth 2023 Oct 20;11:e49677. [doi: 10.2196/49677] [Medline: 37933185]
22. Azimi S, Fernando C, Estai M, Patel J, Silva D, Tennant M. Experience of primary caregivers in utilising an mHealth application for remote dental screening in preschool children. Aust Health Rev 2023 Oct;47(5):545-552. [doi: 10.1071/AH23110] [Medline: 37580061]

XSL•FO
RenderX

23. Hartshorn JE, Nair RU. Dental innovations which will influence the oral health care of baby boomers. Spec Care Dentist 2023;43(3):359-369. [doi: 10.1111/scd.12835] [Medline: 36782274]
24. Menhadji P, Patel R, Asimakopoulou K, et al. Patients' and dentists' perceptions of tele-dentistry at the time of COVID-19. A questionnaire-based study. J Dent 2021 Oct;113:103782. [doi: 10.1016/j.jdent.2021.103782] [Medline: 34400252]
25. Moore R, Brødsgaard I. Dentists' perceived stress and its relation to perceptions about anxious patients. Community Dent Oral Epidemiol 2001 Feb;29(1):73-80. [Medline: 11153566]
26. Hill KB, Hainsworth JM, Burke FJ, Fairbrother KJ. Evaluation of dentists' perceived needs regarding treatment of the anxious patient. Br Dent J 2008 Apr 26;204(8):E13. [doi: 10.1038/sj.bdj.2008.318] [Medline: 18425075]
27. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
28. Scauso MS. Interpretivism: definitions, trends, and emerging paths. In: Oxford Research Encyclopedia of International Studies: Oxford University Press; 2020.
29. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020 May 5;15(5):e0232076. [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
31. Klingberg A, Sawe HR, Hammar U, Wallis LA, Hasselberg M. m-Health for burn injury consultations in a low-resource setting: an acceptability study among health care providers. Telemed J E Health 2020 Apr;26(4):395-405. [doi: 10.1089/tmj.2019.0048] [Medline: 31161967]
32. Compeau DR, Higgins CA. Computer self-efficacy: development of a measure and initial test. MIS Q 1995 Jun 1;19(2):189-211. [doi: 10.2307/249688]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Naeem M, Ozuem W, Howell K, Ranfagni S. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. Int J Qual Methods 2023 Oct;22. [doi: 10.1177/16094069231205789]
35. Charmaz K. Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis: SAGE Publications; 2006.
36. Lincoln YS, Guba EG. Naturalistic Inquiry: SAGE Publications; 1985.
37. Scheerman JF, Qari AH, Varenne B, et al. A systematic umbrella review of the effects of teledentistry on costs and oral-health outcomes. Int J Environ Res Public Health 2024 Mar 28;21(4):407. [doi: 10.3390/ijerph21040407] [Medline: 38673320]
38. Briggie AM, Hilsenroth MJ, Conway F, Muran JC, Jackson JM. Patient comfort with audio or video recording of their psychotherapy sessions: relation to symptomatology, treatment refusal, duration, and outcome. Prof Psychol Res Pract 2016;47(1):66-76. [doi: 10.1037/a0040063]
39. Xiao J, Kopycka-Kedzierawski D, Ragusa P, et al. Acceptance and usability of an innovative mDentistry eHygiene model amid the COVID-19 pandemic within the US National Dental Practice-Based Research Network: mixed methods study. JMIR Hum Factors 2023 Aug 18;10:e45418. [doi: 10.2196/45418] [Medline: 37594795]
40. Kaushik R, Rapaka R. A patient-centered perspectives and future directions in AI-powered teledentistry. Discoveries (Craiova) 2024 Dec 31;12(4):e199. [doi: 10.15190/d.2024.18] [Medline: 40109877]
41. Jawad LA. Security and privacy in digital healthcare systems: challenges and mitigation strategies. Abhigyan 2024;42:23-31. [doi: 10.1177/09702385241233073]
42. Aly NM, El Kashlan MK, Giraudeau N, El Tantawi M. Children's acceptance of intraoral cameras and smartphones for caries detection: a randomized controlled clinical trial. Int J Paediatr Dent 2025 Jul;35(4):792-799. [doi: 10.1111/ipd.13291] [Medline: 39737579]
43. Levin L, Day PF, Hicks L, et al. International Association of Dental Traumatology guidelines for the management of traumatic dental injuries: general introduction. Dent Traumatol 2020 Aug;36(4):309-313. [doi: 10.1111/edt.12574] [Medline: 32472740]
44. Nikielska-Sekula K. "Have you just taken a picture of me?": theoretical and ethical implications of the use of researcher-produced photography in studying migrant minorities. In: Nikielska-Sekula K, Desille A, editors. Visual Methodology in Migration Studies: New Possibilities, Theoretical Implications, and Ethical Questions: Springer Nature; 2021:31-49. [doi: 10.1007/978-3-030-67608-7_2]
45. Waseh S, Dicker AP. Telemedicine training in undergraduate medical education: mixed-methods review. JMIR Med Educ 2019 Apr 8;5(1):e12515. [doi: 10.2196/12515] [Medline: 30958269]
46. Knutson TJ, Komolsevin R, Chatiketu P, Smith V. A cross-cultural comparison of Thai and US American rhetorical sensitivity: implications for intercultural communication effectiveness. Int J Intercult Relat 2003 Feb;27(1):63-78. [doi: 10.1016/S0147-1767(02)00060-3]

Abbreviations

mHealth: mobile health

XSL•FO
RenderX

TAM: technology acceptance model

Edited by K Gierend; submitted 23.Jul.2025; peer-reviewed by C McLeod, K Simmons, S Tungare; revised version received 19.Dec.2025; accepted 07.Mar.2026; published 07.Apr.2026.

Please cite as:
Shenouda M, Huang B, Schultz EC, Estai M, Ranjitkar S, Louie JP, Pungchanchaikul P
Assessing Patient Discomfort in Smartphone-Based Teledentistry From the Perspective of Dental Professionals: Qualitative Interview Study
Online J Public Health Inform 2026;18:e81163
URL: https://ojphi.jmir.org/2026/1/e81163
doi:10.2196/81163

© Margaret Shenouda, Boyen Huang, Emily C Schultz, Mohamed Estai, Sarbin Ranjitkar, Jeff P Louie, Patimaporn Pungchanchaikul. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 7.Apr.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Online Health Information–Seeking Among Older Adults and Predictors of Use, Motivations, and Barriers in the Context of Healthy Aging: Cross-Sectional Study

Yves Bachofner[1], MA; Alexander Seifert[2], PhD; Samin Sepahniya[1], MA; Carlo Fabian[1], MA

[1]Institute for Social Work and Health, School of Social Work, FHNW University of Applied Sciences and Arts, Riggenbachstrasse 16, Olten, Switzerland
[2]Institute for Integration and Participation, School of Social Work, FHNW University of Applied Sciences and Arts, Olten, Switzerland

Corresponding Author:
Yves Bachofner, MA
Institute for Social Work and Health, School of Social Work, FHNW University of Applied Sciences and Arts, Riggenbachstrasse 16, Olten, Switzerland

Abstract

Background: Considering the rapid digital transformation, older adults are increasingly relying on online health information–seeking (OHIS) to support healthy aging. However, disparities in their digital competence levels (the ability to effectively use digital tools) and health literacy (the ability to access, understand, appraise, and apply health information) may influence engagement in OHIS.

Objective: This paper examines the prevalence of OHIS among older adults in Switzerland and identifies their motivations, barriers, and predictors of use. The objective is to determine key factors that promote or hinder OHIS use among older internet users.

Methods: A cross-sectional survey was conducted with 1261 internet users aged 60 years and older living in Switzerland (mean age 70.1, SD 7.3 years; 539/1261, 42.7% female). Descriptive analyses and hierarchical binary logistic regression models were used.

Results: Overall, 77.6% (969/1248) of participants engaged in OHIS in their everyday lives. Subjective health status, internet use frequency, trust in online health information (OHI), and digital competence level significantly influenced OHIS use. Participants reporting good to very good health were less likely to engage in OHIS compared to those in poorer health (odds ratio [OR] 0.496, 95% CI 0.307-0.801; *P*=.004). Higher likelihood of OHIS use was associated with (almost) daily versus less frequent internet use (OR 1.550, 95% CI 1.011-2.376; *P*=.04), general trust versus distrust in OHI (OR 5.784, 95% CI 4.044-8.272; *P*<.001), and advanced versus low digital competence (OR 3.108, 95% CI 1.385-6.975; *P*=.006); health literacy was not a significant predictor of OHIS use (OR 0.912, 95% CI 0.393-2.117; *P*=.83, excellent vs deficient [reference]). Among OHIS users (n=969), the most common frequently indicated motivation for use (672/969, 69.3%) was to gain a better understanding of health conditions. Among nonusers (n=279), the most frequently indicated barriers were difficulties in assessing the credibility of information (159/279, 57%), distrust in the effectiveness of information provided (129/279, 46.2%), and concerns about dubious providers or spam (93/279, 33.3%).

Conclusions: Digital competence, frequent internet use, and trust in OHI are critical for OHIS engagement among older adults. Programs to strengthen digital competencies in later life and initiatives to enhance the credibility of online health resources are essential to reduce digital disparities and support healthy aging. Notably, health literacy did not emerge as a significant factor in OHIS use, but digital competence did, suggesting that digital competence is most critical to OHIS use.

(Online J Public Health Inform 2026;18:e77557) doi:10.2196/77557

KEYWORDS

online health information seeking; healthy aging; digital competence; older adults; health literacy; aging; cross-sectional study; internet; Switzerland

Introduction

Background

With a growing older population, aging presents significant health policy and societal challenges. In response, the World Health Organization's (WHO) "Healthy Aging" [1] framework promotes well-being in later life, emphasizing that functional ability can be maintained despite health challenges. This requires physical and cognitive capacity alongside supportive physical, social, and digital environments [2]. To cope with everyday life, digital competence must increasingly be considered since

XSL•FO
RenderX

digital competence not only is needed for using modern technologies but also enables digital access to health information. The rapid digital transformation, driven by modern information and communication technologies (eg, internet and smartphones), is reshaping knowledge dissemination [3]. While digital solutions enhance quality of life, health, and independence, older adults still use them less than younger groups [2,4]. This digital divide extends beyond access to include disparities in digital competence and use [5]. Indeed, many older adults face challenges due to limited digital competence. Effective digital health promotion requires both access and competencies, highlighting the critical role of digital and health literacy in using digital health services [6].

Online Health Information Seeking Among Older Adults

Digital access is increasingly seen as a key solution for overcoming barriers to obtaining timely health information for older adults [4]. Online health information seeking (OHIS) offers a fast and convenient way to obtain qualitative health-related information but poses challenges due to limited digital competence. Older adults may struggle with navigating sources, formulating queries, and evaluating information and misinformation [7]. Despite greater health concerns, they engage in OHIS less than younger generations, partly due to age-related impairments and digital competence gaps and also because a considerable share of older adults remains offline or does not use internet-enabled devices in the first place. However, even those who use OHIS can benefit from improved access to health information, supporting healthy aging goals [8-10].

Research Questions and Hypotheses

Despite attempts by previous studies [10] to identify the determinants of OHIS in general, the prevalence, motivations, barriers, and predictors of OHIS among older internet users (hereafter referred to as “onliners”) remain largely unclear [7,8,10]. This underscores the need for further investigation to address these gaps.

The aim of this study was to examine the prevalence, motivations, and barriers of OHIS among older onliners in Switzerland and to identify key predictors of OHIS use. Specifically, this study addressed the following research questions: (1) What proportion of onliners aged 60 years and older use OHIS? (2) What are the key determinants of OHIS use? (3) What are the motivations and barriers related to OHIS use?

Regarding the key determinants of OHIS, we proposed hypothesis 1, which assumed that sociodemographic and health-related factors influenced the likelihood of OHIS use. Specifically, we expected that female participants [7], younger individuals (aged 60 - 69 years) [4,11], and participants with higher education levels [12], better financial resources [13], and urban (or intermediate) residency [14] were significantly more likely to use OHIS compared to their counterparts. Regarding health-related factors, we assumed that self-reported health status and the number of medical treatments were associated with OHIS use. While existing evidence was mixed, we expected that individuals with poorer self-reported health statuses [15] and those with more medical treatments [16] in the past year were more likely to use OHIS. Hypothesis 2 assumed that behavioral and attitudinal factors—particularly the frequency of internet use and trust in online health information (OHI)—significantly predicted OHIS use. Specifically, individuals who used the internet daily [16] and those who expressed at least some level of trust in OHI [12,17] were expected to have a greater likelihood of engaging in OHIS. Hypothesis 3 assumed that individual competencies played a critical role in OHIS use. Specifically, higher levels of digital competence [18] and health literacy [19] were expected to increase the probability of OHIS use.

Methods

Study Design and Participants

We conducted a cross-sectional survey within the “Regional Health Promotion in an Age-Friendly Digital World” project with individuals aged 60 years and older living in private households across Switzerland. Participants were sampled by using a stratified random sampling approach using official address data from the Swiss Federal Statistical Office in combination with an additional sampling from the private address provider AZ Direct. Surveys were carried out by Demo Scope AG, an external Swiss pooling provider.

A total of 8311 individuals were invited by mail to participate in the survey, which was available in the 3 official languages of Switzerland (German, French, and Italian). Of these, 1367 (16.4% response rate) completed the survey between June 27 and August 20, 2024, either online (computer-assisted web interviewing: n=1237) or in paper format (paper-and-pencil interviewing: n=130). Incomplete or invalid responses were excluded through rigorous data cleaning, resulting in 1325 valid questionnaires. Of these, 1261 (95.2%) respondents were classified as onliners. For the analyses, we included only the onliners because they had met the basic access requirement for OHIS use.

The questionnaire was developed based on insights from our systematic review [10] and the workshop (n=11) with older adults, family caregivers, and professionals working at the interface of age and health.

Ethical Considerations

The Ethics Committee Northwest and Central Switzerland (Req-2023 - 00727) reviewed this study and determined that it does not fall under the Human Research Act (Art.2). The survey did not collect sensitive health-related personal data, responses were fully anonymized, and participants provided informed consent at the beginning of the survey. No compensation was provided to participants. As such, authorization from the ethics committee was not required.

Measures

The dependent variable, OHIS, was measured via a single item: “In a typical week, how many days do you use websites for getting health-related information?” The question was adapted with minor modifications from the digital health literacy survey instrument developed by the Health Literacy Survey 2019

XSL•FO
RenderX

(HLS19) Consortium of the WHO Action Network on Measuring Population and Organizational Health Literacy [20]. Response options included "more than once per day," "once a day," "4 - 6 days per week," "1 - 3 days per week," "less than once per week," "I don't use it, but it's interesting," and "I don't use it, and I'm not interested in it, either." For analysis, responses indicating any frequency of use ("More than once per day" to "Less than once per week") were recoded as users, while responses indicating no use were recoded as nonusers, resulting in a binary variable (use or nonuse); this approach followed established methods in prior research on OHIS [21].

To explain OHIS use, a range of sociodemographic, health-related, and individual competence factors was considered. Sociodemographic variables included sex (female or male), age group (60 - 69, 70 - 79, and 80 - 100 years), residence location (rural, intermediate, and urban), living arrangement (living alone or with others), education level (compulsory education, secondary education, and tertiary education), and financial situation. The financial situation was assessed through a question adapted from the Swiss Survey on Income and Living Conditions, asking participants how difficult it was for their household to make ends meet with their available income, with responses categorized into "very difficult to rather difficult," "rather simple," and "easy to very easy" [22].

Subjective health status was measured by asking participants to rate their general health, with responses dichotomized afterward into "very poor to mediocre" and "good to very good" categories. To assess the number of medical treatments, participants were asked how often they had received medical treatment (including from general practitioners but excluding dentists) in the previous 12 months. The number of treatments ranged from 0 to 90 (mean 7.28, SD 12.68) and was dichotomized into "below the mean value (of the sample)" and "above the mean value (of the sample)." Both measures were adapted from the Swiss Federal Statistical Office Health Survey [23].

Frequency of internet use was measured by asking how often participants used the internet, with responses dichotomized into "(almost) daily use" and "less than (almost) daily use." Trust in OHI was assessed using participants' responses when asked how trustworthy they found health information from the internet, using a question adapted from Link and Baumann [12], with responses categorized as "rather or very trustworthy, or both trustworthy and not" versus "rather or not at all trustworthy."

Health literacy, defined as the competencies to access, understand, appraise, and apply health information in order to make judgments and take decisions in health care, disease prevention, and health promotion, was assessed using the validated HLS19-Q12 instrument developed by the HLS19 Consortium of the WHO Action Network on Measuring Population and Organizational Health Literacy and categorized into "deficient," "problematic," "sufficient," and "excellent" levels [24]. Digital competence, defined as the ability to use digital technologies in a critical, collaborative, and creative way, was measured using the DigCompSAT tool developed by Clifford et al [25], which was adapted for this study following the approach of Weinhold et al [26] and translated into German, French, and Italian by Stürz et al [27]. The overall score was divided into 4 levels: "low," "basic," "intermediate," and "advanced."

Additionally, OHIS users were asked about their motivations for and nonusers about their barriers to using OHIS, both assessed through multiple response options. The specific response categories for motivations are presented in Table S2 in Multimedia Appendix 1; categories for barriers are in Table S3 in Multimedia Appendix 1.

Analytical Strategy

Statistical analyses were performed using SPSS (version 28; IBM Corp). Descriptive analyses comparing OHIS users (n=969) and nonusers (n=279) and their stated motivations and barriers were conducted using chi-square tests (*P* values) and Cramér *V* (effect size) to assess associations between categorical variables. To identify predictors of OHIS use, a binary logistic regression was performed, allowing for the multivariate analysis of sociodemographic, health-related, and individual competence factors.

Results

Sociodemographic Characteristics of the Sample

The final study sample consisted of 1261 internet users aged 60 years and older, of whom 57.3% (722/1261) were male (Table 1). A total of 52.8% (666/1261) were aged 60 - 69 years, with the overall mean age being 70.1 (SD 7.3) years. Most participants lived in urban areas (718/1261, 57%), and the majority did not live alone (936/1228, 76.2%). Regarding educational attainment, 57.5% (714/1242) had completed secondary school, and 36.9% (458/1242) held a tertiary degree.

XSL•FO
RenderX

Table . Sample characteristics (N=1261) among participants aged 60 years and older who use the internet (onliners, aged 60 years and older).

	Sample, n (%)
(Registered) sex	
Female	539 (42.7)
Male	722 (57.3)
Age groups (years)	
60 - 69	666 (52.8)
70 - 79	438 (34.7)
80 - 100	157 (12.5)
Residence location	
Rural	265 (21)
Intermediate	278 (22)
Urban	718 (57)
Living arrangement	
Living alone	292 (23.8)
Not alone	936 (76.2)
No information	33
Education	
Compulsory	70 (5.6)
Secondary school II	714 (57.5)
Tertiary level	458 (36.9)
No information	19
Financial situation	
Very difficult to rather difficult	236 (19.5)
Rather simple	334 (27.6)
Easy to very easy	639 (52.9)
No information	52
Subjective health status	
Very poor to mediocre	294 (23.5)
Good to very good	959 (76.5)
No information	8
Number of medical treatments	
Below the mean value	910 (75.9)
Above the mean value	289 (24.1)
No information	62

Financial situation was described as easy to very easy by 52.9% (639/1209), rather simple by 27.6% (n=334), and rather to very difficult by 19.5% (n=236). Most participants reported good to very good health (959/1253, 76.5%). The number of medical treatments in the previous 12 months ranged from 0 to 90; 75.9% (910/1199) were below and 24.1% (289/1199) above the sample mean (mean 7.28, SD 12.68). Table 1 provides the sample characteristics.

Use of OHIS

Among onliners aged 60 years and older, 77.6% (969/1248) reported engaging in OHIS, while 22.4% (279/1248) did not. OHIS use was more frequent among female users (429/534, 80.3%) than male users (540/714, 75.6%), and this difference was statistically significant. Age differences were also significant, with the highest OHIS use among participants aged 60 - 69 years (531/658, 80.7%), compared to 70 - 79 years (320/434, 73.7%) and 80 years and older (118/156, 75.6%). Education level showed a significant association with OHIS use, with the highest use among those with tertiary education

XSL•FO
RenderX

(384/455, 84.4%) compared to secondary (523/707, 74%) and compulsory schooling (49/68, 72.1%).

No significant bivariate associations were observed for residence location, living arrangement, financial situation, subjective health status, or number of medical treatments. See Table 2 for full distributions.

Table . Characteristics of online health information seeking (OHIS) users (n=969) and nonusers (n=279) among participants aged 60 years and older who use the internet (onliners, aged 60 years and older).

	OHIS user (n=969), n (%)	OHIS nonuser (n=279), n (%)	Cramér *V*[a]	*P* value
(Registered) sex			0.056	.048
Female	429 (80.3)[b]	105 (19.7)		
Male	540 (75.6)	174 (24.4)		
Age groups (years)			0.079	.02
60 - 69	531 (80.7)	127 (19.3)		
70 - 79	320 (73.7)	114 (26.3)		
80 - 100	118 (75.6)	38 (24.4)		
Residence location			0.043	.31
Rural	201 (76.4)	62 (23.6)		
Intermediate	206 (74.9)	69 (25.1)		
Urban	562 (79.2)	148 (20.8)		
Living arrangement			0.032	.27
Living alone	219 (75.5)	71 (24.5)		
Not alone	728 (78.6)	198 (21.4)		
Education			0.123	<.001
Compulsory	49 (72.1)	19 (27.9)		
Secondary school II	523 (74)	184 (26)		
Tertiary level	384 (84.4)	71 (15.6)		
Financial situation			0.048	.25
Very difficult to rather difficult	174 (74)	61 (26)		
Rather simple	261 (78.9)	70 (21.1)		
Easy to very easy	501 (79.1)	132 (20.9)		
Subjective health status			0.047	.10
Very poor to mediocre	237 (81.2)	55 (18.8)		
Good to very good	726 (76.6)	222 (23.4)		
Number of medical treatments			0.001	.98
Below the mean value	701 (77.5)	203 (22.5)		
Above the mean value	222 (77.6)	64 (22.4)		

[a]Reported Cramér *V* values with corresponding *P* values indicate the strength and significance of group differences.

[b]Percentages are calculated within subgroups (users vs nonusers).

Predictors of OHIS

To identify significant predictors of OHIS, 3 hierarchical binary logistic regression models were conducted. These models sequentially examined the effects of sociodemographic (sex, age, education, financial situation, residence location, and living arrangement) and health-related (subjective health and number of medical treatments) factors (model 1), internet use and trust in OHI (model 2), and individual health literacy and digital competence (model 3; Table 3).

XSL•FO
RenderX

Table . Binary logistic regression[a] models predicting online health information seeking (OHIS) use among onliners aged 60 years and older (n=1043) across sociodemographic and health-related factors, internet use and online health information (OHI) trust, and individual competence[b].

Predictors	Model 1: sociodemographic and health-related factors		Model 2: model 1 factors plus internet use and OHI trust		Model 3: model 2 factors plus digital competence and health literacy	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
(Registered) sex						
Male	Reference	Reference	Reference	Reference	Reference	Reference
Female	1.369 (0.981-1.912)	.06	1.295 (0.902-1.860)	.16	1.409 (0.972-2.043)	.07
Age groups (years)						
60 - 69	Reference	Reference	Reference	Reference	Reference	Reference
70 - 79	0.696 (0.498-0.972)	.03	0.757 (0.527-1.088)	.13	0.782 (0.540-1.132)	.19
80 - 100	0.989 (0.424-1.122)	.13	0.790 (0.465-1.343)	.38	0.884 (0.512-1.524)	.66
Residence location						
Rural	Reference	Reference	Reference	Reference	Reference	Reference
Intermediate	1.020 (0.641-1.621)	.94	1.032 (0.625-1.706)	.90	1.010 (0.607-1.681)	.97
Urban	1.094 (0.740-1.618)	.65	0.998 (0.652-1.528)	.99	0.983 (0.638-1.514)	.94
Living arrangement						
Living alone	Reference	Reference	Reference	Reference	Reference	Reference
Not alone	1.271 (0.876-1.844)	.21	1.325 (0.886-1.982)	.17	1.319 (0.877-1.982)	.18
Education						
Compulsory	Reference	Reference	Reference	Reference	Reference	Reference
Secondary school II	1.115 (0.566-2.196)	.75	0.943 (0.442-2.009)	.88	0.748 (0.346-1.619)	.46
Tertiary level	1.994 (0.964-4.1259)	.06	1.353 (0.601-3.050)	.47	0.996 (0.432-2.293)	.99
Financial situation						
Very difficult to rather difficult	Reference	Reference	Reference	Reference	Reference	Reference
Rather simple	1.356 (0.860-2.138)	.19	1.332 (0.813-2.182)	.26	1.310 (0.794-2.162)	.29
Easy to very easy	1.394 (0.917-2.120)	.12	1.381 (0.873-2.186)	.17	1.322 (0.824-2.121)	.25
Subjective health status						
Very poor to mediocre	Reference	Reference	Reference	Reference	Reference	Reference
Good to very good	0.537 (0.344-0.837)	.006	0.505 (0.315-0.811)	.005	0.496 (0.307-0.801)	.004
Number of medical treatments						
Below the mean value	Reference	Reference	Reference	Reference	Reference	Reference
Above the mean value	0.780 (0.522-1.167)	.23	0.753 (0.488-1.162)	.20	0.774 (0.501-1.198)	.25
Internet use						
Less than (almost) daily	__[d]	—	Reference	Reference	Reference	Reference
(Almost) daily internet use	—	—	1.970 (1.321-2.937)	<.001	1.550 (1.011-2.376)	.04
Trust in OHI						

XSL•FO
RenderX

Predictors	Model 1: sociodemographic and health-related factors		Model 2: model 1 factors plus internet use and OHI trust		Model 3: model 2 factors plus digital competence and health literacy	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Rather or not at all trustworthy	—	—	Reference	Reference	Reference	Reference
OHI are rather or very trustworthy, or both trustworthy and not	—	—	6.026 (4.252-8.542)	<.001	5.784 (4.044-8.272)	<.001
Health literacy (HLS19-Q12)						
Deficient	—	—	—	—	Reference	Reference
Problematic	—	—	—	—	0.733 (0.400-1.346)	.32
Sufficient	—	—	—	—	0.669 (0.349-1.282)	.23
Excellent	—	—	—	—	0.912 (0.393-2.117)	.83
Digital competence (DigCompSAT)						
Low	—	—	—	—	Reference	Reference
Basic	—	—	—	—	1.811 (0.990-3.316)	.05
Intermediate	—	—	—	—	2.660 (1.467-4.824)	.001
Advanced	—	—	—	—	3.108 (1.385-6.975)	.006

[a]Dependent variable: user OHIS=1, nonuser OHIS=0. For detailed statistical values (CIs), please refer to Table S1 in Multimedia Appendix 1.

[b]Model fit: model 1: Nagelkerke R^2=0.045; χ^2_{12}=30.2; *P*=.003; model 2: Nagelkerke R^2=0.217; χ^2_{14}=154.3; *P*<.001; and model 3: Nagelkerke R^2=0.234; χ^2_{20}=167.4; *P*<.001.

[c]OR: odds ratio.

[d]The predictor was not included in the respective model.

Model 1 (Nagelkerke R^2=0.045; χ^2_{12}=30.2; *P*=.003) was statistically significant and revealed that only age was a significant predictor within the sociodemographic variables. Participants aged 70 - 79 years were significantly less likely to use OHIS compared to those aged 60 - 69 years (odds ratio [OR] 0.696, 95% CI 0.498-0.972; *P*=.03). Notably, no significant difference was observed between participants aged 80 - 100 years and those aged 60 - 69 years (OR 0.989, 95% CI 0.424-1.122; *P*=.13). In contrast, other sociodemographic factors that were significant in the bivariate analysis—sex and education level—did not retain significance in the multivariate model. Besides age, subjective health was also a significant predictor. Participants who rated their health as good to very good were less likely to use OHIS compared to those with poor to mediocre health (OR 0.537, 95% CI 0.344-0.837; *P*=.006). Conversely, the number of medical treatments in the previous year showed no significant association with OHIS engagement (OR 0.780, 95% CI 0.522-1.167; *P*=.23). These results provide mixed support for hypothesis 1.

Model 2 (Nagelkerke R^2=0.217; χ^2_{14}=154.3; *P*<.001) introduced internet use frequency and trust in OHI as predictors. The analysis revealed that both factors were significant predictors of OHIS use, providing full support for hypothesis 2. Participants who reported using the internet (almost) daily were nearly twice as likely to use OHIS compared to those who used it less frequently (OR 1.970, 95% CI 1.321 - 2.937; *P*<.001). Additionally, participants who perceived OHI as rather or very trustworthy, or both trustworthy and not, were over 6 times more likely to use OHIS than those who distrusted OHI (OR 6.026, 95% CI 4.252 - 8.542; *P*<.001). Notably, the previously significant effect of age became nonsignificant after including these 2 model 2 variables (OR 0.757, 95% CI 0.527-1.088; *P*=.13).

Model 3 (Nagelkerke R^2=0.234; χ^2_{20}=167.4; *P*<.001) added health literacy and digital competence to the analysis. Compared to adults with low digital competence levels, those with intermediate competence were more than twice as likely to use OHIS (OR 2.660, 95% CI 1.467 - 4.824; *P*=.001), and those with advanced competence were over 3 times more likely (OR 3.108, 95% CI 1.385 - 6.975; *P*=.006) to use OHIS. In contrast, health literacy was not a significant predictor. Additionally, subjective health status, daily internet use, and trust in OHI continued to be significant predictors in model 3.

The model's explanatory power increased with each step, as indicated by the rising Nagelkerke R^2, from 0.045 in model 1 to 0.234 in model 3. This progression highlights how the inclusion of internet use, trust in OHI, and digital competence substantially improved the model's ability to predict OHIS use.

Motivations for OHIS

Among the 969 OHIS users, the most commonly indicated reason for use was to gain a better understanding of certain health conditions or illnesses (672/969, 69.3%), followed by learning about medications and their possible side effects (538/969, 55.5%) and searching for treatment options or therapies for specific health problems (528/969; 54.5%; Table

XSL•FO
RenderX

4). Additionally, searching for alternative or complementary medical approaches (424/969, 43.8%) and seeking information out of general interest (402/969, 41.5%) were notable motivations. Fewer participants indicated using OHIS to obtain a second opinion (180/969, 18.6%) or for other reasons (9/969, 0.9%; eg, assisting family members and searching for information when health professionals are unavailable).

Table . Motivations for engaging in online health information seeking (OHIS) among OHIS users (n=969) within the onliner population (aged 60 years and older), including chi-square tests for sex and age differences[ab].

Motivation (multiple response options)	Total, n (%)	Male, n (%)	Female, n (%)	Chi-square test for differences in sex, *P* value	60 - 69 years, n (%)	70 - 79 years, n (%)	80 - 100 years, n (%)	Chi-square test for differences in age, *P* value
Understanding health conditions	672 (69.3)	372 (68.9)	300 (69.9)	.73	370 (69.7)	216 (67.5)	86 (72.9)	.54
Medications and side effects	538 (55.5)	284 (52.6)	254 (59.2)	.04	268 (50.5)	197 (61.6)	73 (61.9)	.002
Treatment options or therapies	528 (54.5)	262 (48.5)	266 (62)	<.001	274 (51.6)	190 (59.4)	64 (54.2)	.09
Alternative or complementary medicine	424 (43.8)	192 (35.6)	232 (54.1)	<.001	239 (45)	139 (43.4)	46 (39)	.49
Just out of interest	402 (41.5)	233 (43.1)	169 (39.4)	.24	241 (45.4)	112 (35)	49 (41.5)	.01
Second opinion	180 (18.6)	118 (21.9)	62 (14.5)	.003	89 (16.8)	63 (19.7)	28 (23.7)	.18
Other reasons	9 (0.9)	8 (1.5)	1 (0.2)	N/A[c]	4 (0.8)	3 (0.9)	2 (1.7)	N/A

[a]Detailed effect sizes (Cramér *V*) and full answer options from the survey are reported in Table S2 in Multimedia Appendix 1.

[b]Sorted by total.

[c]N/A indicates that no calculation was performed because cells had a frequency of fewer than 5.

Sex differences were significant for several motivations. Female participants were more likely than male participants to search for information on treatment options or therapies (266/429, 62% vs 262/540, 48.5%), alternative or complementary medical approaches (232/429, 54.1% vs 192/540, 35.6%), and medications and side effects (254/429, 59.2% vs 284/540, 52.6%). Conversely, male participants were more inclined to search for a second opinion (118/540, 21.9% vs 62/429, 14.5%).

Significant age-related differences also emerged. Older participants, particularly those aged 70 - 79 (197/320, 61.6%) and 80 - 100 years (73/118, 61.9%), were more likely to seek information about medications and side effects compared to the 60 - to 69-year age group (268/531, 50.5%). In contrast, younger participants (aged 60-69 years) were more likely to search for OHI out of general interest (241/531, 45.4%) than older groups.

Barriers to OHIS

The most commonly indicated barrier to use among OHIS nonusers was difficulty assessing the credibility of information (159/279, 57%), followed by distrust in the effectiveness of the information provided (129/279, 46.2%), concerns about dubious providers or the risk of spam and advertising (93/279, 33.3%), lack of experience with searching for information on the internet (87/279, 31.2%), and challenges related to technical or difficult-to-understand language in health information (46/279, 16.5%; Table 5). Fewer participants indicated barriers such as lack of support in using digital services (20/279, 7.2%), negative past experiences with online searches (17/279, 6.1%), physical limitations when using digital devices (10/279, 3.6%), and other reasons (51/279, 18.3%, eg, outdated or unclear publication dates and lack of personal interest in health information). Sex- or age-related differences did not attain statistical significance for any of the barriers.

XSL•FO
RenderX

Table . Barriers to engaging in online health information seeking (OHIS) among OHIS nonusers (n=279) within the onliner population (60 years and older), including chi-square tests for sex and age differences[ab].

Barriers (multiple response options)	Total, n (%)	Male, n (%)	Female, n (%)	Chi-square test for differences in sex, *P* value	60 - 69 years, n (%)	70 - 79 years, n (%)	80 - 100 years, n (%)	Chi-square test for differences in age, *P* value
Credibility	159 (57)	96 (55.2)	63 (60)	.43	71 (55.9)	64 (56.1)	24 (63.2)	.71
Distrust	129 (46.2)	84 (48.3)	45 (42.9)	.38	61 (48)	51 (44.7)	17 (44.7)	.86
Dubious offers	93 (33.3)	60 (34.5)	33 (31.4)	.60	47 (37)	39 (34.2)	7 (18.4)	.10
Lack of experience	87 (31.2)	56 (32.2)	31 (29.5)	.64	32 (25.2)	38 (33.3)	17 (44.7)	.06
Technical language	46 (16.5)	31 (17.8)	15 (14.3)	.44	20 (15.7)	20 (17.5)	6 (15.8)	.93
Lack of support	20 (7.2)	12 (6.9)	8 (7.6)	.82	8 (6.3)	8 (7)	4 (10.5)	N/A[c]
Negative experiences	17 (6.1)	12 (6.9)	5 (4.8)	.47	10 (7.9)	4 (3.5)	3 (7.9)	N/A
Physical limitations	10 (3.6)	4 (2.3)	6 (5.7)	N/A	3 (2.4)	5 (4.4)	2 (5.3)	.59
Other reasons	51 (18.3)	30 (17.2)	21 (20)	.56	20 (15.7)	23 (20.2)	8 (21.1)	.60

[a]Detailed effect sizes (Cramér *V*) and full answer options from the survey are reported in Table S3 in Multimedia Appendix 1.

[b]Sorted by total.

[c]N/A indicates that no calculation was performed because cells had a frequency of fewer than 5.

Discussion

Principal Findings

The study findings revealed that OHIS occurred widely among older adults in this demographic, with 77.6% (n=969) of older onliners using OHIS. This aligns with prior research demonstrating high engagement with digital health resources among older adults [28]. Notably, no significant difference in OHIS engagement was found between individuals aged 80 - 100 years and the younger age groups, although a drop in use was observed in the 70 - to 79-year age group compared to the 60 - to 69-year group. This suggests that the oldest age group may have adapted to digital tools similarly to younger older adults [4]. One potential explanation for this negligible discrepancy may be that the younger age group (60-69 years) was more inclined to experiment with technology and explore digital tools, consequently resulting in higher OHIS use. In contrast, the oldest group (80-100 years) may be more predisposed to seek information online for health reasons [29]. Furthermore, this study revealed a marginally elevated propensity among female participants to use OHIS, aligning with the extant literature suggesting that female participants exhibit a heightened propensity to proactively seek health-related information [7].

Education emerged as a significant predictor of OHIS use. Individuals with tertiary education were more likely to seek health information online, supporting the theory of the digital divide, where higher education correlates with better digital competence and greater access to online resources [5].

In the multivariate analysis, the effects of education, sex, and age lost statistical significance. This suggests that, while these sociodemographic factors may initially appear associated with OHIS use, their explanatory power diminishes when health, behavioral, and competence-related variables, such as subjective health status, digital competence, and trust in OHI, are considered. This pattern aligns with previous findings that highlight the centrality of these more proximal determinants [21]. This highlights the importance of broader structural and individual determinants in shaping OHIS use.

Markedly, individuals with poorer self-reported health statuses were more likely to use OHIS, supporting findings that health concerns drive proactive information seeking [30]. However, the number of medical treatments was not associated with OHIS engagement, suggesting that health care use alone does not motivate OHIS. Instead, sufficient information from health care providers may reduce the need for additional online searches, while other providers may encourage OHIS use [16].

The predictive role of digital competence was shown within our analyses; people with higher levels of digital competence were more often within the group of OHIS users. A higher level of digital competence can facilitate the ability to search for OHI, while those with low competence levels remained disengaged, despite internet access, underscoring that mere access is insufficient for effective use [6,18].

Moreover, regular use of the internet also predicted OHIS use and can be regarded as a behavioral indicator of technological familiarity, thereby further supporting the application of OHIS. However, digital competence encompasses a more extensive ability to effectively engage with digital tools across various contexts.

Contrary to the findings of other studies, health literacy was not a significant predictor of OHIS use in this research [7,19]. This

XSL•FO
RenderX

suggests that, while individuals with lower health literacy may face challenges in comprehending and critically evaluating health information, these difficulties do not necessarily prevent them from OHIS engagement. The ease of access and widespread availability of OHI may encourage use regardless of comprehension levels. However, this raises concerns about the potential risk of misinterpretation or reliance on misleading information, particularly among those with lower health literacy levels. This highlights that OHIS primarily reflects the act of searching rather than the quality of comprehension or application, a finding consistent with Wang et al [21], who emphasized that instrumental factors, such as utility and trust, are far stronger predictors of OHIS than psychological or cognitive abilities related to processing health information. As a result, individuals with lower health literacy may still use OHIS without necessarily deriving meaningful health benefits. This underscores the need for integrated strategies that strengthen both digital competence and health literacy to ensure that access to information translates into informed decision-making and improved health outcomes.

Of the variables included, trust in OHI proved to be the strongest predictor of OHIS use. Participants who perceived OHI as trustworthy were significantly more likely to engage in OHIS, underscoring the central role that perceived credibility plays in online health behaviors. This finding aligns with prior research, which has consistently shown that trust is a key determinant in digital health use [7,21,31,32].

Conversely, a lack of trust in OHI was among the barriers most frequently cited by nonusers. This distrust often stems from concerns about misinformation, unreliable sources, and commercial influences [31]. In line with previous studies, respondents expressed apprehension regarding the credibility of online health resources, which aligns with findings from Sbaffi and Rowley [33], who emphasized that website design, intrusive advertisements, and complex language negatively affect the perceived trustworthiness of OHI.

Importantly, sex and age differences indicated distinct information-seeking patterns, with female participants more focused on treatment-related topics and alternative medicine and male participants more likely to seek second opinions, while younger participants demonstrated a broader, more general interest in health-related content compared to older age groups. Therefore, digital health information should always consider the different audiences and, if necessary, tailor its content to specific audiences.

Implications for Practice and Policy

Enhancing digital competence through targeted training could improve OHIS use, especially among older adults with low digital competence levels [7]. Public health campaigns should build trust in OHI by promoting credible and user-friendly digital health platforms. Addressing individual capabilities and improving the quality of digital health information can help bridge gaps in OHIS use [30]. As highlighted by Jacob et al [34], the effectiveness of digital health interventions depends not only on providing information but also on ensuring user trust through privacy, security, and credibility. For offline individuals, the challenge lies in gaining access to digital resources. Expanding digital infrastructures and providing accessible training are essential first steps toward enabling digital engagement [10]. However, reliable offline health information (eg, flyers and brochures from government health organizations) must continue to be available to meet the needs of those who do not engage with digital platforms.

Limitations

This study has several limitations. As it focuses on Switzerland alone, the generalizability of our findings to other contexts may be limited. The cross-sectional design prevents time comparisons and, therefore, causal conclusions about factors influencing OHIS use. Hence, future longitudinal studies should investigate factors that influence changes in OHIS use over time. Self-reported data, such as subjective health, may introduce recall or social desirability bias, potentially affecting the accuracy of responses. Additionally, the content and quality of the accessed health information were not assessed, limiting insights into the variance of the individual user profiles.

Conclusions

This paper highlights the significant correlation of subjective health status, digital competence, daily internet use, and trust in OHI with OHIS use among older adults. Health literacy and sociodemographic characteristics showed no significant correlation when examined alongside other factors. Addressing digital competence and enhancing trust in OHI are essential for reducing digital inequalities and empowering older adults to manage their health more actively, thereby promoting healthy aging.

Acknowledgments

The authors would like to thank the older adults who participated in the study.

Funding

This study was supported by the Swiss National Science Foundation as part of the project “Regional Health Promotion in an Age-Friendly Digital World” (grant IZCOZ0_213366). The funder had no role in designing the study or interpreting the data.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: YB, AS
Data curation: YB, AS
Formal analysis: YB, AS
Funding acquisition: AS, CF
Investigation: YB, AS
Project administration: CF
Visualization: YB
Writing—original draft: YB
Writing—review and editing: YB, AS, SS, CF

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental tables for detailed statistical values.
[DOCX File, 60 KB - ojphi_v18i1e77557_app1.docx]

References

1. Decade of healthy ageing: baseline report. World Health Organization. 2020. URL: https://iris.who.int/handle/10665/338677 [accessed 2024-08-05]
2. Cotten SR. Technologies and aging: understanding use, impacts, and future needs. In: Ferraro KF, Carr D, editors. Handbook of Aging and the Social Sciences, 9th edition: Academic Press; 2021:373-392. [doi: 10.1016/B978-0-12-815970-5.00023-1]
3. Fang ML, Canham SL, Battersby L, Sixsmith J, Wada M, Sixsmith A. Exploring privilege in the digital divide: implications for theory, policy, and practice. Gerontologist 2019 Jan 9;59(1):e1-e15. [doi: 10.1093/geront/gny037] [Medline: 29750241]
4. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
5. van Dijk JAGM. Digital divide research, achievements and shortcomings. Poetics (Amst) 2006 Aug;34(4-5):221-235. [doi: 10.1016/j.poetic.2006.05.004]
6. Cornejo Müller A, Wachtler B, Lampert T. Digital divide—soziale unterschiede in der nutzung digitaler gesundheitsangebote [Digital divide—social differences in the use of digital health services]. Bundesgesundheitsbl 2020 Feb;63(2):185-191. [doi: 10.1007/s00103-019-03081-y]
7. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790. [doi: 10.2196/34790] [Medline: 35171099]
8. Shutsko A. Country-specific digital inequalities in older people's online health information seeking in Europe: impact of socio-demographic and socio-economic factors. J Inf Manag 2022;10(4):38-52. [doi: 10.1633/JISTaP.2022.10.4.4]
9. Seifert A, Cotten SR. Use of information and communication technologies among older adults: usage differences, health-related impacts, and future needs. In: Rohlinger DA, Sobieraj S, Rohlinger DA, Sobieraj S, editors. The Oxford Handbook of Digital Media Sociology: Oxford University Press; 2022:291-315. [doi: 10.1093/oxfordhb/9780197510636.013.12]
10. Bachofner Y, Seifert A, Sepahniya S, Fabian C. Exploring online health information seeking and sharing among older adults: a mini-review about acceptance, potentials, and barriers. Front Digit Health 2024;6:1336430. [doi: 10.3389/fdgth.2024.1336430] [Medline: 38313862]
11. Hargittai E, Piper AM, Morris MR. From internet access to internet skills: digital inequality among older adults. Univ Access Inf Soc 2019 Nov;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
12. Link E, Baumann E. Nutzung von gesundheitsinformationen im internet: personenbezogene und motivationale einflussfaktoren [Use of health information on the internet: personal and motivational factors]. Bundesgesundheitsbl 2020 Jun;63(6):681-689. [doi: 10.1007/s00103-020-03144-5]
13. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
14. Zhang Y, Lin Z, Li X, Xiaoming T, Zhou Y, Zhang X. Factors affecting ICT use in health communication among the older population in Jiangsu, China. Libri 2019;69(1):41-53. [doi: 10.1515/libri-2018-0103]
15. Weber W, Reinhardt A, Rossmann C. Lifestyle segmentation to explain the online health information-seeking behavior of older adults: representative telephone survey. J Med Internet Res 2020 Jun 12;22(6):e15099. [doi: 10.2196/15099] [Medline: 32530433]
16. Choi NG, DiNitto DM, Lee OE, Choi BY. Internet and health information technology use and psychological distress among older adults with self-reported vision impairment: case-control study. J Med Internet Res 2020 Jun 3;22(6):e17294. [doi: 10.2196/17294] [Medline: 32490851]

XSL•FO
RenderX

17. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011 Feb 16;13(1):e19. [doi: 10.2196/jmir.1552] [Medline: 21324832]
18. Welch V, Petkovic J, Pardo Pardo J, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75. [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
19. Rockmann R, Gewald H, Brune P. Gesundheitsbezogene IT-nutzung im altersübergang—ursachen und auswirkungen individueller differenzen [Health-related IT use in the transition to old age—causes and effects of individual differences]. In: Schneider W, Stadelbacher S, editors. Der Altersübergang Als Neuarrangement von Arbeit Und Leben: Kooperative Dienstleistungen Für Das Alter(n) in Vielfalt Wiesbaden: Springer Fachmedien; 2019:177-198. [doi: 10.1007/978-3-658-21974-1_8]
20. The HLS19 Consortium of the WHO Action Network M-POHL. The HLS19-DIGI instruments for measuring digital health literacy factsheet. : Austrian National Public Health Institute; 2024 URL: https://m-pohl.net/sites/m-pohl.net/files/2024-03/Factsheet_p%20HLS19-DIGI.pdf [accessed 2025-12-31]
21. Wang X, Shi J, Kong H. Online health information seeking: a review and meta-analysis. Health Commun 2021 Sep;36(10):1163-1175. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
22. Bundesamt für Statistik (BFS). Haushaltsfragebogen. erhebung über die einkommen und lebensbedingungen (SILC) [household questionnaire. survey on income and living conditions (SILC)]. : Bundesamt für Statistik; 2024 URL: https://www.bfs.admin.ch/bfs/en/home/statistics/economic-social-situation-population/surveys/silc.html [accessed 2026-01-01]
23. Bundesamt für Statistik (BFS). Schweizerische gesundheitsbefragung 2022—telefonischer und schriftlicher fragebogen [swiss health survey 2022—telephone and written questionnaire]. : Eidgenössisches Departement des Innern EDI, Bundesamt für Statistik BFS, Abteilung Gesundheit und Soziales; 2023 URL: https://www.bfs.admin.ch/bfs/en.assetdetail.28965074.html [accessed 2026-01-01]
24. Pelikan JM, Link T, Straßmayr C, et al. Measuring comprehensive, general health literacy in the general adult population: the development and validation of the HLS19-Q12 instrument in seventeen countries. Int J Environ Res Public Health 2022 Oct 29;19(21):14129. [doi: 10.3390/ijerph192114129] [Medline: 36361025]
25. Clifford I, Kluzer S, Troia S, Jakobsone M. DigCompSat. European Commission. 2020. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC123226 [accessed 2025-12-27]
26. Weinhold N, Barczik K, Jokisch M, Doh M, Göbl L. Einführung in die messung digitaler kompetenzen im alter: bestandsaufnahme und erste ableitungen zur erfassung der digitalen kompetenz bei älteren menschen [Introduction to measuring digital skills in older adults: inventory and initial conclusions on assessing digital skills in older adults]. MedienPädagogik Sep 2023:581-618. [doi: 10.21240/mpaed/jb20/2023.09.22.X]
27. Stürz RA, Stumpf C, Schlude A, Mendel U. Das bidt-digitalbarometer. international [the bidt digital barometer international]. : Bidt—Bayerisches Forschungsinstitut Für Digitale Transformation; 2023 URL: https://publikationen.badw.de/en/049378486/049378486.pdf [accessed 2026-01-01]
28. Oh YS, Lim J. Patient-provider communication and online health information seeking among a sample of US older adults. J Health Commun 2021 Oct 3;26(10):708-716. [doi: 10.1080/10810730.2021.1998846] [Medline: 34743676]
29. Hunsaker A, Hargittai E. A review of internet use among older adults. New Media Soc 2018 Oct;20(10):3937-3954. [doi: 10.1177/1461444818787348]
30. Di Novi C, Kovacic M, Orso CE. Online health information seeking behavior, healthcare access, and health status during exceptional times. J Econ Behav Organ 2024 Apr;220:675-690. [doi: 10.1016/j.jebo.2024.02.032] [Medline: 38628501]
31. Waterworth S, Honey M. On-line health seeking activity of older adults: an integrative review of the literature. Geriatr Nurs 2018;39(3):310-317. [doi: 10.1016/j.gerinurse.2017.10.016] [Medline: 29198622]
32. Crespo-Gonzalez C, Benrimoj SI, Frommer M, Dineen-Griffin S. Navigating online health information: Insights into consumer influence and decision-making strategies—an overview of reviews. Digit Health 2024;10:20552076241286815. [doi: 10.1177/20552076241286815] [Medline: 39493637]
33. 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. [doi: 10.2196/jmir.7579] [Medline: 28630033]
34. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]

Abbreviations

HLS19: Health Literacy Survey 2019
OHI: online health information
OHIS: online health information seeking
OR: odds ratio
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Tsehay; submitted 15.May.2025; peer-reviewed by H Siamian, N Gordon; revised version received 10.Nov.2025; accepted 14.Nov.2025; published 06.Jan.2026.

Please cite as:
Bachofner Y, Seifert A, Sepahniya S, Fabian C
Online Health Information–Seeking Among Older Adults and Predictors of Use, Motivations, and Barriers in the Context of Healthy Aging: Cross-Sectional Study
Online J Public Health Inform 2026;18:e77557
URL: https://ojphi.jmir.org/2026/1/e77557
doi:10.2196/77557

© Yves Bachofner, Alexander Seifert, Samin Sepahniya, Carlo Fabian. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 6.Jan.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patient Comfort With Smartphone-Based and Intraoral Camera Imaging in Teledentistry: Observational Cross-Sectional Study

Martha B Lucas[1], BS; Boyen Huang[1], DDS, MHA, PhD; Ryan HL Ip[2], BSc, PhD; Angela M Hastings[1], DDS; Emily C Schultz[3], RDH, MS; Yao-Yi Chiang[4], PhD; Jeffrey P Louie[5], MD

[1]Department of Primary Dental Care, School of Dentistry, University of Minnesota, Minneapolis, MN, United States
[2]Department of Mathematical Sciences, Auckland University of Technology, Auckland, Auckland, New Zealand
[3]Department of Dental Education, College of Allied Health and Nursing, Minnesota State University, Mankato, Mankato, MN, United States
[4]Department of Computer Science & Engineering, College of Science and Engineering, University of Minnesota, Minneapolis, MN, United States
[5]Department of Pediatrics, Medical School, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Boyen Huang, DDS, MHA, PhD
Department of Primary Dental Care
School of Dentistry
University of Minnesota
515 Delaware Street SE
Minneapolis, MN, 55455
United States
Phone: 1 612 624 7848
Email: huan2321@umn.edu

Abstract

Background: Smartphone-based imaging has shown promise in teledentistry; yet, patient comfort during its use compared with intraoral cameras remains underexplored.

Objective: This study aimed to compare self-perceived comfort during oral evaluations using smartphones and intraoral cameras in a teledentistry setting.

Methods: A convenience sample of 192 participants aged 6 years and older completed a survey and underwent 2 procedures: trained researchers captured 3 photographs of the anterior teeth with a Samsung smartphone, and the same teeth were examined with a MouthWatch intraoral camera. Comfort was rated by participants before and after each procedure using a pictorial 7-point Likert scale, and data were analyzed with nonparametric tests.

Results: Comfort decreased in 15% (29/192) of the participants following intraoral camera examination (P=.001), while no significant change was observed after smartphone imaging (P=.52). The difference in comfort change between smartphone imaging and intraoral camera examination was marginal (P=.05). Among the small subgroup of youths (aged 6-19 years; n=13, 6.8%), 38% (5/13) reported a reduction in comfort after smartphone imaging, compared with 10% (18/179) among adults (P=.02). In addition, significantly more participants without tooth wear reported a reduction in comfort after smartphone imaging (5/12, 42%), compared with 10% (18/180) among those with tooth wear (P=.02).

Conclusions: Overall, the findings indicate generally comparable comfort levels between the 2 approaches, with a marginal difference favoring smartphone imaging. While the potential for broader use of smartphone-based imaging in a teledentistry setting still requires further research, targeted education and aligned institutional and insurance policies may enable its scalable, sustainable integration that enhances patient experience.

(Online J Public Health Inform 2026;18:e95563) doi:10.2196/95563

KEYWORDS

mobile health; mHealth; patient comfort; tooth structure loss; crown fracture; caries; tooth wear

XSL•FO
RenderX

Introduction

Teledentistry uses digital technologies to deliver synchronous and asynchronous oral health care and dental assessments to patients who are not physically present with dental professionals. Initially developed to support dental consultations for US troops, teledentistry helped improve access to care and reduce travel burdens within military settings [1]. In addition to expanding access, teledentistry has been promoted as a strategy to reduce oral health inequalities and lessen the financial impact associated with traditional dental care [2]. Teledentistry models vary, with some enabling real-time interactions between providers and patients through video consultations and others relying on photographs captured for clinical evaluation [3]. In many cases, intraoral cameras are used to obtain images of the oral cavity for remote assessment [4-6]. More recently, teledentistry has expanded into mobile health apps, in which patients or caregivers use smartphone cameras to take photographs that are later reviewed by dental professionals. Studies have demonstrated the feasibility of remote assessment of dental caries [3], traumatic dental injuries [7], and dental restoration quality [8] using smartphone-based imaging. Similarly, the feasibility and reliability of teledentistry assessments for dental caries [9] and oral potentially malignant disorders [10] using images captured with intraoral cameras have been confirmed. Despite this potential, access to in-person dental care and diagnosis remains limited for many individuals, especially those in low-income, uninsured, and remote communities who face a disproportionately high burden of oral disease [11]. Because traditional diagnostic visits can be costly and time-intensive, alternative approaches are needed. Teledentistry provides a means of remote assessment and patient triage, improving access to care while reducing demand on conventional dental settings [12].

Although evidence supports the diagnostic accuracy of teledentistry, barriers to adoption persist. While many dental professionals express support for teledentistry [13,14], patient acceptance varies considerably. Reported barriers include the absence of physical contact during examinations [15], perceived psychological, financial, and performance risks [16], limited access to digital devices or limited digital literacy [15], concerns about privacy and confidentiality [17], and discomfort with capturing intraoral photographs [15]. Social influences also play a role, with greater acceptance observed when family and peers endorse teledentistry [16], and patient anxiety during procedures further affects compliance and outcomes [18]. Together, these factors underscore that patient comfort and acceptance are critical for the successful integration of teledentistry into routine care.

While both smartphones and intraoral cameras are increasingly used for the detection of dental caries and oral cancer [4-6], most comparative research has centered on diagnostic accuracy [5,6]. Far less attention has been given to patient-centered outcomes, particularly comfort and acceptance. To date, only one study has examined children's acceptance of both modalities during caries screening [4], highlighting a notable gap in understanding how adolescents and adults experience these tools in terms of emotional and physical comfort. In this context, comfort refers to a state of ease marked by the absence of distress and the presence of safety, positivity, and situational acceptance [19]. Research in other clinical fields has documented patient comfort with mobile imaging technologies, such as smartphone-based photography in neurology [20] and portable otoscope telemedicine devices in otology [21]. Within teledentistry, however, only one recent study has evaluated patient comfort with smartphone-based imaging from the perspective of dental clinicians, and it did not directly capture patients' own interpretations of comfort [17].

To address this gap, this study aimed to assess prospective patients' comfort levels during examinations conducted with an intraoral camera versus those involving smartphone-captured photographs. The goal was to generate evidence that could guide strategies for expanding the use of teledentistry among the general population and, more specifically, to support broader access to dental care in remote and underserved communities. Accordingly, this study tested whether patient comfort differed between the 2 imaging modalities.

Methods

Ethical Considerations

This study was approved by the institutional review board of the University of Minnesota (STUDY00022279). Informed consent was obtained from all adult participants, and parental consent was obtained for those younger than 18 years. All study personnel followed institutional review board–approved procedures for data access, confidentiality, and privacy. All data were anonymized and deidentified prior to analysis. Participants received a spin toothbrush as compensation after providing consent and completing the study procedures.

Study Design and Setting

This study adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline for cross-sectional studies [22]. The observational cross-sectional study was conducted in Minnesota, United States, in 2024, as part of a multisite machine learning project designed to advance teledentistry technology and optimize user experience for image-based assessments of tooth structure loss (TSL), encompassing crown fractures, caries, and tooth wear [23]. Researchers used the Minnesota State Fair as a community-based setting to recruit prospective dental patients who may or may not have had immediate dental care needs. Recruitment and data collection took place inside the University of Minnesota's Driven to Discover building, a designated research facility located on the fairgrounds that provided space for fairgoers and researchers to come together for study participation and community engagement. The Minnesota State Fair is an annual ticketed event featuring agricultural exhibitions, art, food, entertainment, and carnival rides. Every year, it attracts approximately 2 million fairgoers from across the country over 12 days from late August to early September [24], offering a diverse pool of potential participants.

Participants

To support a separate machine learning analysis using the same dataset, approximately 150 dental photos were needed for each

XSL•FO
RenderX

TSL condition [25], contributing to a cumulative target of 600 participants. Recruitment was planned across multiple sites, with at least 200 participants enrolled at each, including the Minnesota State Fair and 2 international collaborating locations. Eligibility criteria required participants to be at least 6 years old, exhibit one or more signs of TSL, and be able to communicate in English.

Recruitment followed a standardized procedure. Researchers were stationed at the entrance of the Driven to Discover building during designated data collection hours (9 AM-8 PM on 3 separate days). Passersby were visually screened for potential eligibility based on age and the presence of visible signs of TSL. Researchers used a brief, scripted verbal explanation to invite eligible individuals to participate. Additional participants self-referred after viewing a signboard indicating that participation included receiving a spin toothbrush. Participants with identifiable features in or around the mouth (eg, piercings, birthmarks, tattoos, or jewelry) were excluded. Given the open and recreational nature of the fair environment, convenience sampling was used.

Devices and Materials

Dental photos were taken using the camera function of a Samsung Galaxy S20 FE 5G smartphone, without any image editing, and then securely uploaded to the Box cloud storage system. These smartphones were purchased new for our previous study [7], restricted to password-protected institutional Wi-Fi (eduroam), prohibited from installing additional apps, and stored and managed by a key researcher. Individual user accounts and passwords were required for Box. MouthWatch intraoral cameras were used to assess the participants' anterior teeth and to complete a dental chart, which also operated independently of linked software. A survey tool, designed to collect demographic information and evaluate comfort levels, was administered using iPads (Apple Inc). Additionally, a dental chart was accessed through the Notability app on the iPad and later transferred to a spreadsheet for auditing and analysis.

Procedure

All researchers were trained and calibrated before participant recruitment and data collection. Each participant completed a self-administered demographic questionnaire. Using a pictorial version of a 7-point Likert scale [26], participants rated their comfort levels before and after each imaging method. Examinations were conducted in a fixed sequence, beginning with the intraoral camera and followed by smartphone imaging with flash. Due to the recreational and high-throughput nature of the Minnesota State Fair setting, procedural simplicity was required in the affiliated research facilities; therefore, the order of modalities was not randomized or counterbalanced.

Consistent with prior publications using smartphone imaging in dentistry, 3 standardized photographs were captured for each participant (Figure 1) [7]. Images were taken with the designated smartphone under natural light with the flash feature activated. Each participant was seated upright in a chair, and the researcher stood directly in front during image capture. All images were immediately uploaded via a secure wireless network to the Box cloud system. An experienced teledentist subsequently performed a remote assessment, charting tooth-by-tooth conditions, including uncomplicated and complicated crown fractures, caries, tooth wear, and teeth that were missing, restored, sound, or partially erupted.

Figure 1. Example dental photographs captured during the study using a Samsung Galaxy S20 FE 5G smartphone in a cross-sectional teledentistry evaluation of tooth structure loss among fairgoers at the 2024 Minnesota State Fair (United States): (A) mandibular occlusal view, (B) frontal view, and (C) maxillary occlusal view. For infection-control purposes, participants retracted their lips with their own fingers during image capture. The images are shown with the participant's consent.

Data Analysis

Data entry and statistical analysis were performed using Excel (Microsoft Office LTSC Professional Plus 2024) and R (version 4.3.3; R Foundation for Statistical Computing), respectively. Counties of residence were categorized as either "metropolitan" or "nonmetropolitan" according to the 2023 Rural-Urban Continuum Codes [27].

Due to the subjective nature of the comfort levels, the "before and after differences," defined as the comfort level after a photo is taken minus that before a photo is taken for the same participant, were used as the outcome variables to reduce the effects of within-participant variability. The changes in comfort levels after each device were assessed among all participants, as well as among different subgroups formed by potential confounding variables such as demographics and dental conditions. The changes in comfort levels across all participants were analyzed using the Wilcoxon signed-rank test. The changes in comfort levels between different subgroups were analyzed using the Wilcoxon rank-sum test (when there were 2 subgroups) and Kruskal-Wallis test (when there are 3 or more subgroups). The use of nonparametric methods was appropriate due to the nonnormal nature of the data collected. The level of significance was set at 5%.

XSL•FO
RenderX

Results

226 fairgoers consented to participate. After excluding participants with missing data, a total of 192 (85%) participants were included in the final sample. These included 181 (94.3%) residing in Minnesota and 11 (5.7%) from other US states. The teledentist's remote assessment confirmed TSL conditions (crown fractures, caries, or tooth wear) in 190 (99%) participants. Table 1 presents the frequency distribution of the final sample.

XSL•FO
RenderX

Table 1. Frequency distribution of demographic and dental characteristics and associated changes in comfort levels among 192 participants in an observational cross-sectional teledentistry study conducted at the 2024 Minnesota State Fair (United States), including *P* values for comfort change after smartphone imaging, after intraoral camera examination, and between modalities.

	Frequency, n (%)	Comfort level change after smartphone imaging, *P* value	Comfort level change after intraoral camera examination, *P* value	Difference in comfort level change between smartphone imaging and intraoral camera examination, *P* value
Demographics				
Gender		.69	.19	.32
Woman	98 (51)			
Man	92 (48)			
Nonbinary	2 (1.0)			
Race		.30	.68	.93
Black, Indigenous, and People of Color	30 (15.6)			
White	162 (84.4)			
Age (y)[a]		.25	.20	.44
6-19	13 (6.8)			
20-29	25 (13)			
30-39	17 (8.9)			
40-49	24 (12.5)			
50-59	36 (18.7)			
60-69	51 (26.6)			
70-84	26 (13.5)			
Age group		.02[b]	.32	.08
Adolescents	13 (6.8)			
Adults	179 (93.2)			
Area of residence		.76	.08	.10
Metropolitan	174 (90.6)			
Nonmetropolitan	18 (9.4)			
Dental conditions				
Caries		.18	.26	.80
Caries	32 (16.7)			
No caries	160 (83.3)			
Crown fracture		.86	.53	.34
Crown fracture	158 (82.3)			
No crown fracture	34 (17.7)			
Tooth wear		.02[b]	.80	.03[b]
Tooth wear	180 (93.8)			
No tooth wear	12 (6.2)			
Restoration		.85	.15	0.2
Restoration	164 (85.4)			
No restoration	28 (14.6)			
Overall[c]	192 (100)	.52	.001[b]	.05

[a]Data analysis based on a Kruskal-Wallis test.

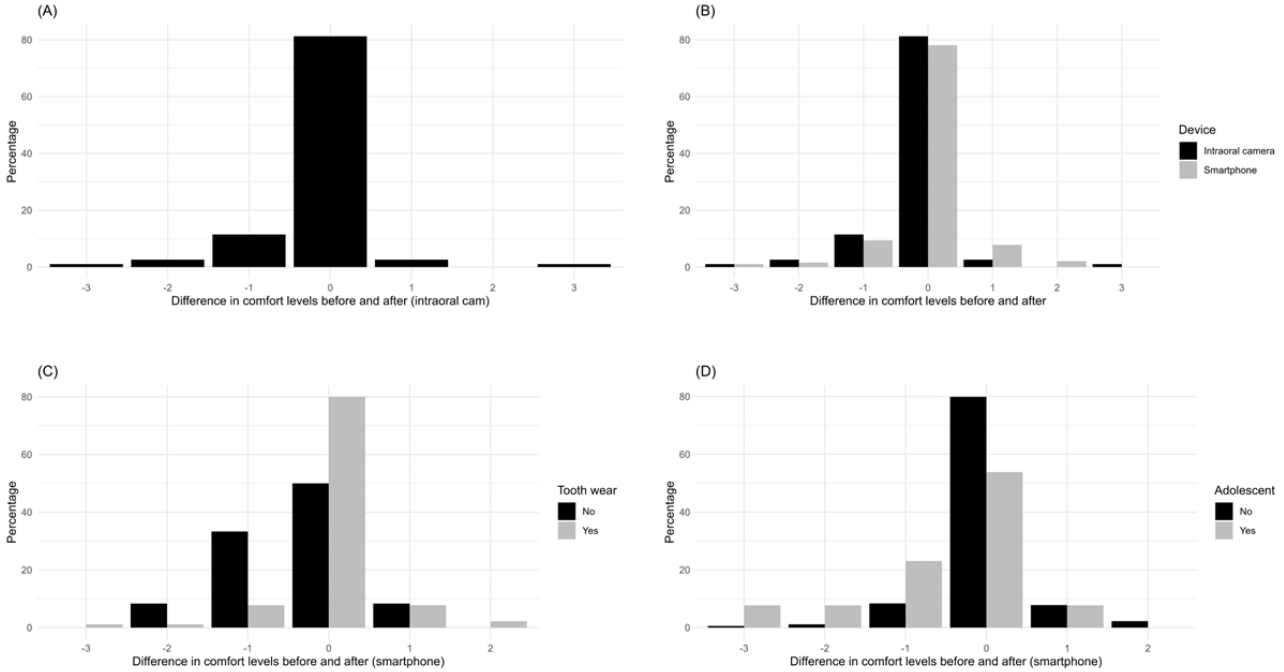
XSL•FO
RenderX

[b]$P<.05$.

[c]Data analysis based on a Wilcoxon signed-rank test.

More participants reported a decrease in comfort levels than an increase following the intraoral camera examination (P=.001; Figure 2A). The result of the Wilcoxon signed-rank test reflects the asymmetry of the distribution. Of the 192 participants, 29 (15.1%) reported a reduction in comfort after the intraoral camera procedure, while only 7 (3.6%) indicated an increase in comfort. No significant change in comfort was observed after smartphone imaging (P=.52). When comparing the 2 modalities, the change in comfort levels before and after the procedure showed only a marginal difference between smartphone imaging and intraoral camera examination (P=.05; Figure 2B).

Figure 2. Changes in participant comfort levels before and after imaging procedures in an observational cross-sectional study of teledentistry conducted at the 2024 Minnesota State Fair (United States): (A) comfort change after intraoral camera examination, (B) comparison of comfort change between smartphone imaging and intraoral camera examination, (C) comfort change after smartphone imaging by tooth wear status, and (D) comfort change after smartphone imaging by age group.



Participants without tooth wear experienced a greater decrease in comfort levels after smartphone imaging compared with those with tooth wear (P=.02; Figure 2C). Among the 12 participants without tooth wear, 5 (41.7%) reported a reduction in comfort levels, while only 10% (18/180) of the participants with tooth wear reported the same. It should be noted that among the 12 participants without tooth wear, 4 (33.3%) were younger than 20 years, 1 (8.3%) had caries, 8 (66.7%) had uncomplicated crown fractures, and 2 (16.7%) had complicated crown fractures. Note that these categories are not mutually exclusive and therefore the sum of the reported counts is greater than 12. Moreover, children and adolescents (aged 6-19 years) experienced a more pronounced reduction in comfort after smartphone use than adults (P=.02; Figure 2D). Among the 13 children and adolescents, 5 (38.5%) indicated a reduction in comfort, while only 10.1% (18/179) of participants aged 20 years or older reported the same.

No significant differences in comfort levels were observed by gender, metropolitan or nonmetropolitan residence, or other oral conditions (P≥.08).

Discussion

Interpretation of Findings

This study aimed to compare patient comfort during dental evaluations conducted with smartphone-based imaging and intraoral camera examination. Consistent with this objective, the findings indicated that comfort decreased following intraoral camera use, whereas smartphone imaging did not produce a significant change. When comparing the 2 modalities, smartphone imaging demonstrated comparable comfort, with a trend toward greater comfort relative to intraoral camera examination. These results directly addressed the study aim and contributed new evidence to the limited literature on patient-centered outcomes in teledentistry.

The consistent comfort associated with smartphone imaging may be due to factors such as familiarity, patient friendliness, simplicity, and portability, as suggested by Aly et al [4]. Given the widespread use of smartphones for communication, education, and recreation [28], individuals may feel more at ease, less intimidated, and more comfortable when photographed with such devices [4]. Similar preferences have been observed in other areas of dentistry where less invasive digital technologies, such as digital impression systems, have replaced

XSL•FO
RenderX

traditional impression techniques [29]. In those contexts, patients have favored digital methods due to increased efficiency, reduced chair time, and reduced discomfort associated with prolonged mouth opening [30]. Similarly, in esthetic dentistry, digital 3D modeling has been preferred over conventional approaches for its shorter treatment duration, enhanced comfort, and reliability [31]. Although Aly et al [4] noted that smartphone imaging elicited lower fear and better behavior among pediatric patients, they also found that intraoral camera procedures were completed more quickly [4]. However, patients undergoing intraoral camera examinations experienced difficulty maintaining mouth opening [4]. Thus, perceived familiarity, ease of use, and reduced duration of mouth opening during smartphone imaging may matter more than its brevity. Future investigations are warranted to examine how these advantages, individually or in combination, affect patient comfort.

The present study identified an association between age and patient comfort levels, with children and adolescents exhibiting a greater reduction in comfort following smartphone imaging compared with adults. This pattern contrasts with a prior study reporting more favorable behavior in children examined using a smartphone compared with an intraoral camera [4]. In our sample, a smaller proportion of adults than children reported decreased comfort after smartphone imaging, and this age-related finding should be interpreted cautiously given the small number of youth participants (n=13) and its exploratory nature. This observation may reflect the well-documented inverse association between age and dental fear, which is more prevalent among younger populations [32]. Fear of dental visits and avoidance behaviors are especially common in children and adolescents, who may have difficulty understanding the purpose of dental examinations, particularly within a research context. The introduction of smartphone technology in a diagnostic dental setting may have further compounded their discomfort. Together, elevated dental anxiety and limited familiarity with smartphone-based imaging in a clinical environment may explain the greater decline in comfort among younger participants. To better support pediatric patients during smartphone-based teledentistry, clinical communication should incorporate friendly, interactive explanations of the procedures and include opportunities for children to view the images captured of their own teeth.

Subgroup analyses revealed that participants without tooth wear experienced a greater decline in comfort after smartphone imaging than those with tooth wear. No associations were observed between changes in comfort and other dental conditions, including caries, uncomplicated crown fractures, or complicated crown fractures. As documented in the literature, the prevalence of tooth wear increases with age [33], suggesting that participants without tooth wear were more likely to be younger. This was reflected in the present study's distribution: 4 (33.3%) of the 12 participants without tooth wear were younger than 20 years, markedly higher than the overall proportion of the entire sample (13/192, 6.8%). Given the inverse relationship between age and comfort levels observed in this study, the greater discomfort reported by participants without tooth wear may be confounded by age. Furthermore, most participants without tooth wear (n=12) presented with other TSL signs: 1 with caries, 8 with uncomplicated crown fractures, and 2 with complicated crown fractures. Dental anxiety, often triggered by the discovery of oral disease and the anticipated need for intervention [34], is associated with avoidance of dental visits and poorer health outcomes [35]. Increased distress among patients with untreated dental conditions and a heightened perceived need for care may explain the discomfort reported by those without tooth wear during smartphone imaging [18]. Because the subgroup without tooth wear was small, these findings should be considered exploratory and validated in future research with larger and more balanced samples.

No differences in comfort levels were observed based on participants' gender, race, or area of residence. Although prior research has not specifically examined comfort responses to teledentistry imaging modalities across these demographic groups, earlier studies have identified associations between demographic factors and dental anxiety. For example, female children [36], those of Asian descent [37], or those residing in rural areas have been shown to exhibit higher levels of dental anxiety [36]. Nevertheless, these findings should not be extrapolated to infer similar patterns in comfort with smartphone or intraoral camera imaging. Recent work has instead highlighted factors such as "clear communication, informed consent, and assurances regarding privacy and data security" as key contributors to patient comfort during smartphone-based dental imaging [17]. These considerations point to communication quality as a contextual factor that may shape patient experience across demographic groups.

Limitations

Recruitment at the Minnesota State Fair introduced several limitations to this study. Participation in a large-scale outdoor event required the purchase of an entry ticket and the physical ability to navigate the fairgrounds, which may have excluded individuals with limited financial resources or physical disabilities. Moreover, individuals with more severe forms of TSL, such as deep caries with cellulitis or complicated crown fractures accompanied by bleeding or pain, were unlikely to attend a recreational venue, resulting in probable underrepresentation of these conditions. The study's inclusion criteria focused exclusively on TSL, thereby excluding individuals with other oral health conditions such as periodontal disease, oral cancer, or xerostomia.

The recreational nature of the state fair also required procedural simplicity. Consequently, the order of imaging modalities was fixed rather than randomized, with all participants undergoing intraoral camera examination first, followed by smartphone imaging. This lack of randomization may have introduced an order-effect bias [38], whereby the experience of the first procedure could influence perceived comfort during the second. Although the direction of such an effect cannot be determined from the present data, the potential for order-related influence remains an important methodological limitation. Future studies should randomize or counterbalance the order of procedures to minimize this bias.

This study also used a single smartphone model and a single intraoral camera model, which limits the generalizability of the

XSL•FO
RenderX

findings. Future research should incorporate multiple device types and diverse clinical settings to strengthen external validity.

Some statistically significant results were derived from relatively small subgroup sample sizes. Therefore, caution is required when interpreting these findings, and future studies should aim to recruit larger and more balanced subgroup samples.

Finally, the use of a convenience sample limited the generalizability of the findings. Random sampling was not feasible in the state fair setting, and similar constraints would likely arise in clinical recruitment. To enhance representativeness in future research, sampling across multiple community or school settings may provide a more diverse and generalizable participant pool.

Conclusions

Within the study limitations, these findings suggest that smartphone-based imaging may serve as a patient-friendly alternative to intraoral cameras in both clinical and teledentistry settings. Beyond the immediate comparison of comfort levels, the results indicate that smartphone imaging may warrant further evaluation as a scalable, accessible component of remote dental assessment, particularly in underserved or resource-limited communities. Integrating smartphone-based imaging into routine practice may reduce barriers related to equipment cost, portability, and patient acceptance. To facilitate broader adoption, targeted education for patients and dental professionals, along with supportive institutional and insurance policies, will be essential. Future research should explore implementation strategies, longitudinal patient experiences, and the role of smartphone imaging in comprehensive teledentistry workflows.

Acknowledgments

The authors acknowledge funding support from the Frank W Veden Endowment for Clinical Research, the University of Minnesota School of Dentistry, and the University of Minnesota Medical School Division of Pediatric Emergency Medicine and Child Abuse. The authors also thank the student workers and staff members at the University of Minnesota for their assistance. Part of this work was presented at the 54th General Session and Exhibition of the American Association for Dental, Oral, and Craniofacial Research (AADOCR, 2025). Portions of this manuscript were rephrased with the assistance of Microsoft Copilot, a generative AI tool developed by Microsoft. Microsoft Copilot was used solely for language refinement and did not generate or edit any original scientific content, data, analyses, or interpretations. All AI-assisted text was reviewed and verified by the authors to ensure accuracy, integrity, and the absence of plagiarism.

Data Availability

The deidentified datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

This study was supported by the Frank W Veden Endowment for Clinical Research Fund, the Summer Research Fellowship Program funding for MBL and Faculty Start-up Funding for BH, both from the University of Minnesota School of Dentistry, and the University of Minnesota Medical School Division of Pediatric Emergency Medicine and Child Abuse Fund. The funding sources had no role in the study design and were not involved in its execution, data analysis, interpretation, or the decision to submit the results.

Authors' Contributions

Conceptualization: BH (lead), AMH, ECS, YYC, JPL
Data curation: BH (lead), MBL
Formal analysis: RHLI
Funding acquisition: BH (lead), AMH, ECS, YYC, JPL
Investigation: MBL (equal), BH (equal), AMH (equal)
Methodology: BH (lead), AMH, ECS
Project administration: MBL
Resources: BH (lead), JPL
Supervision: BH
Validation: AMH (lead), MBL, BH
Visualization: RHLI (lead), MBL, BH
Writing—original draft: MBL
Writing—review and editing: BH (lead), RHLI, AMH, ECS, YYC, JPL

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Rocca MA, Kudryk VL, Pajak JC, Morris T. The evolution of a teledentistry system within the Department of Defense. Proc AMIA Symp 1999:921-924 [FREE Full text] [Medline: 10566495]
2. Talla PK, Rodrigues A. What does teledentistry mean? Mapping of current definitions. Digit Health 2024 Nov 25;10:20552076241292681 [FREE Full text] [doi: 10.1177/20552076241292681] [Medline: 39600384]
3. Estai M, Kanagasingam Y, Huang B, Checker H, Steele L, Kruger E, et al. The efficacy of remote screening for dental caries by mid-level dental providers using a mobile teledentistry model. Community Dent Oral Epidemiol 2016 Oct;44(5):435-441. [doi: 10.1111/cdoe.12232] [Medline: 27111291]
4. Aly NM, El Kashlan MK, Giraudeau N, El Tantawi M. Children's acceptance of intraoral cameras and smartphones for caries detection: a randomized controlled clinical trial. Int J Paediatr Dent 2025 Jul;35(4):792-799. [doi: 10.1111/ipd.13291] [Medline: 39737579]
5. Mahrous R, Amer H, Essam W. Accuracy of intraoral camera and smartphone camera in teledetection of dental caries among adults: a diagnostic accuracy study. J Evid Based Dent Pract 2025 Sep;25(3):102160. [doi: 10.1016/j.jebdp.2025.102160] [Medline: 40716829]
6. Nguyen J, Yang S, Melnikova A, Abouakl M, Lin K, Takesh T, et al. Novel approach to improving specialist access in underserved populations with suspicious oral lesions. Curr Oncol 2023 Jan 11;30(1):1046-1053 [FREE Full text] [doi: 10.3390/curroncol30010080] [Medline: 36661729]
7. Huang B, Estai M, Pungchanchaikul P, Quick K, Ranjitkar S, Fashingbauer E, et al. Mobile health assessment of traumatic dental injuries using smartphone-acquired photographs: a multicenter diagnostic accuracy study. Telemed J E Health 2024 Oct;30(10):2592-2600 [FREE Full text] [doi: 10.1089/tmj.2024.0012] [Medline: 38946689]
8. Yazicioglu A, Yildiz Telatar G. Quality assessment of anterior composite restorations: clinical examination versus three digital photographic techniques. Clin Oral Investig 2026 Apr 18;30(5):181. [doi: 10.1007/s00784-026-06870-3] [Medline: 41999411]
9. Tan WK, Chua DR. Parental use and acceptance of an accessible, commercially available intraoral camera for teledentistry in their children. Eur Arch Paediatr Dent 2024 Apr;25(2):237-246. [doi: 10.1007/s40368-024-00878-7] [Medline: 38643420]
10. Vetchaporn S, Rangsri W, Ittichaicharoen J, Rungsiyakull P. Validity and reliability of intraoral camera with fluorescent aids for oral potentially malignant disorders screening in teledentistry. Int J Dent 2021 Oct 29;2021:6814027 [FREE Full text] [doi: 10.1155/2021/6814027] [Medline: 34745263]
11. Northridge ME, Kumar A, Kaur R. Disparities in access to oral health care. Annu Rev Public Health 2020 Apr 02;41:513-535 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094318] [Medline: 31900100]
12. Malpe M, Choudhari SG, Nagtode N, Muntode Gharde P. Beyond the chair: exploring the boundaries of teledentistry. Cureus 2024 Jun 13;16(6):e62286. [doi: 10.7759/cureus.62286] [Medline: 39006674]
13. Alabass SO, Elhadi YA, Saeed Alhussain MM, Siddig RE, Abdalla Elfaki AO, Yousif Elgayli EE, et al. Acceptance and readiness for tele-dentistry among dental professionals amidst conflict in Sudan. BMC Oral Health 2025 Jul 06;25(1):1123 [FREE Full text] [doi: 10.1186/s12903-025-06501-5] [Medline: 40619363]
14. Schultz EC, Huang B, Shenouda M, Estai M, Ranjitkar S, Louie JP, et al. Perspectives of frontline clinicians and end-line users on smartphone-based photography for assessing traumatic dental injuries: focus group interview study and thematic analysis. JMIR Hum Factors 2026 Jan 26;13:e82668 [FREE Full text] [doi: 10.2196/82668] [Medline: 41587769]
15. Tenore G, Podda GM, La Torre G, Rocchetti F, Palaia G, Di Paolo C, et al. Evaluation of patient's perception and acceptance of tele(oral) medicine for care during the COVID-19 pandemic: a cross-sectional pilot survey. Appl Sci 2021 Aug 13;11(16):7443. [doi: 10.3390/app11167443]
16. Bahanan L, Alsharif M. Factors affecting the acceptance of teledentistry determined using the technology acceptance model: a cross-sectional study. Digit Health 2023 Feb 19;9:20552076231158034 [FREE Full text] [doi: 10.1177/20552076231158034] [Medline: 36825078]
17. Shenouda M, Huang B, Schultz EC, Estai M, Ranjitkar S, Louie JP, et al. Assessing patient discomfort in smartphone-based teledentistry from the perspective of dental professionals: qualitative interview study. Online J Public Health Inform 2026 Apr 07;18:e81163 [FREE Full text] [doi: 10.2196/81163] [Medline: 41945764]
18. Armfield JM, Stewart JF, Spencer AJ. The vicious cycle of dental fear: exploring the interplay between oral health, service utilization and dental fear. BMC Oral Health 2007 Jan 14;7:1 [FREE Full text] [doi: 10.1186/1472-6831-7-1] [Medline: 17222356]
19. Wensley C, Botti M, McKillop A, Merry AF. Maximising comfort: how do patients describe the care that matters? A two-stage qualitative descriptive study to develop a quality improvement framework for comfort-related care in inpatient settings. BMJ Open 2020 May 18;10(5):e033336 [FREE Full text] [doi: 10.1136/bmjopen-2019-033336] [Medline: 32430447]
20. He G, Dunn HP, Ahmad KE, Watson E, Henderson A, Tynan D, et al. Fundoscopy use in neurology departments and the utility of smartphone photography: a prospective prevalence and crossover diagnostic accuracy study amongst neurology inpatients. Eur J Neurol 2022 Aug;29(8):2463-2472 [FREE Full text] [doi: 10.1111/ene.15390] [Medline: 35531644]

XSL•FO
RenderX

21. Hesaka N, Nakazawa T, Kakehata S. A portable pen-shaped otoscope for telemedicine and office-based otologic examination: feasibility and patient acceptability. J Clin Med 2026 Jan 27;15(3):1028 [FREE Full text] [doi: 10.3390/jcm15031028] [Medline: 41682709]
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
23. Jang L, Chiang YY, Hastings AM, Pungchanchaikul P, Lucas MB, Schultz EC, et al. OMNI-Dent: towards an accessible and explainable AI framework for automated dental diagnosis. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops. New York, NY: IEEE; 2026:415-424.
24. Attendance. Minnesota State Fair. URL: https://www.mnstatefair.org/about/attendance/ [accessed 2025-10-26]
25. Shahinfar S, Meek P, Falzon G. “How many images do I need?” Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. Ecol Inform 2020 May;57:101085. [doi: 10.1016/j.ecoinf.2020.101085]
26. Williams JM, Murray JJ, Lund CA, Harkiss B, de Franco A. Anxiety in the child dental clinic. J Child Psychol Psychiatry 1985 Mar;26(2):305-310. [doi: 10.1111/j.1469-7610.1985.tb02268.x] [Medline: 3980617]
27. Rural-Urban Continuum Codes. Economic Research Service, U.S. Department of Agriculture. 2025. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes [accessed 2025-10-26]
28. Bientzle M, Restle A, Kimmerle J. Perception of purposeful and recreational smartphone use in physiotherapy: randomized controlled trial. JMIR Mhealth Uhealth 2021 Apr 21;9(4):e25717 [FREE Full text] [doi: 10.2196/25717] [Medline: 33881402]
29. Siqueira R, Galli M, Chen Z, Mendonça G, Meirelles L, Wang HL, et al. Intraoral scanning reduces procedure time and improves patient comfort in fixed prosthodontics and implant dentistry: a systematic review. Clin Oral Investig 2021 Dec;25(12):6517-6531 [FREE Full text] [doi: 10.1007/s00784-021-04157-3] [Medline: 34568955]
30. Yuzbasioglu E, Kurt H, Turunc R, Bilir H. Comparison of digital and conventional impression techniques: evaluation of patients' perception, treatment comfort, effectiveness and clinical outcomes. BMC Oral Health 2014 Jan 30;14:10 [FREE Full text] [doi: 10.1186/1472-6831-14-10] [Medline: 24479892]
31. Mocelin RC, Penteado MM, Pierre FZ, Saraiva AC, Uemura ES, da Silva JM. Assessment of patient and dentist preference between conventional and digital diagnostic waxing. Int J Esthet Dent 2021 Aug 17;16(3):300-309. [Medline: 34319665]
32. Grisolia BM, Dos Santos AP, Dhyppolito IM, Buchanan H, Hill K, Oliveira BH. Prevalence of dental anxiety in children and adolescents globally: a systematic review with meta-analyses. Int J Paediatr Dent 2021 Mar;31(2):168-183. [doi: 10.1111/ipd.12712] [Medline: 33245591]
33. Decup F, Léger S, Lefèvre S, Doméjean S, Grosgogeat B. Risk factors or indicators for dental caries and tooth wear and their relative importance in adults according to age. J Dent 2024 Aug;147:105092. [doi: 10.1016/j.jdent.2024.105092] [Medline: 38821393]
34. Hatch JP, Mohebzad E. The fearful and phobic patient. In: Prevention in Clinical Oral Health Care. St. Louis, MO: Mosby; 2008.
35. Alenezi AA, Aldokhayel HS. The impact of dental fear on the dental attendance behaviors: a retrospective study. J Family Med Prim Care 2022 Oct;11(10):6444-6450 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_1030_22] [Medline: 36618260]
36. Vasakova J, Tichy A, Duskova J, Bradna P, Broukal Z. Dental fear/anxiety in Czech preschoolers: relationship to gender, age, location of residence, dental status and previous treatment experience. Eur J Paediatr Dent 2022 Dec;23(4):327-331 [FREE Full text] [doi: 10.23804/ejpd.2022.23.04.13] [Medline: 36511903]
37. Mahiepala NA, Phan VL, Kieu KD, Koppen JP, Hussain BH, Huang B. Influencing factors of paediatric dental anxiety levels in an undergraduate dental clinic. Eur J Paediatr Dent 2015 Jun;16(2):159-162. [Medline: 26147825]
38. Lucas CP. The order effect: reflections on the validity of multiple test presentations. Psychol Med 1992 Feb;22(1):197-202. [doi: 10.1017/s0033291700032852] [Medline: 1574556]

Abbreviations

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TSL: tooth structure loss

XSL•FO
RenderX

Edited by E Mensah; submitted 17.Mar.2026; peer-reviewed by P Dhawan; comments to author 15.Apr.2026; revised version received 01.Jul.2026; accepted 10.Jul.2026; published 05.Aug.2026.

Please cite as:
Lucas MB, Huang B, Ip RHL, Hastings AM, Schultz EC, Chiang YY, Louie JP
Patient Comfort With Smartphone-Based and Intraoral Camera Imaging in Teledentistry: Observational Cross-Sectional Study
Online J Public Health Inform 2026;18:e95563
URL: https://ojphi.jmir.org/2026/1/e95563
doi:10.2196/95563
PMID:

©Martha B Lucas, Boyen Huang, Ryan HL Ip, Angela M Hastings, Emily C Schultz, Yao-Yi Chiang, Jeffrey P Louie. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 05.Aug.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Tweets Surrounding Pharmaceutical Drug Brands With Top Direct-to-Consumer TV-Advertising Budgets: Social Media Listening Study

Wendy Macias[1*], PhD; María Belén Navarro Castillo[2*], MS

[1]Department of Strategic Communication, Texas Christian University, TCU Box 298065, Fort Worth, TX, United States

[2]AECOM, Dallas, TX, United States

[*]all authors contributed equally

Corresponding Author:
Wendy Macias, PhD
Department of Strategic Communication, Texas Christian University, TCU Box 298065, Fort Worth, TX, United States

Abstract

Background: Direct-to-consumer (DTC) pharmaceutical advertising allocates billions annually in the United States; however, the analysis of conversations on social media about DTC drugs remains sparse. Twitter (subsequently rebranded X) is serving as a forum for pharmaceutical companies, their constituents, and social media health influencers to discuss topics related to DTC drugs with high advertising budgets.

Objective: This study aims to examine user-generated topics discussed for the highest-budgeted DTC pharmaceutical drugs and to identify the agenda-setting themes and uses and gratifications that emerged.

Methods: This social media listening study used Brandwatch to analyze Twitter conversations (August 2021-August 2023) surrounding top-budgeted DTC brands from 2020 to 2022. A dataset of 44,700 mentions from 26,800 unique authors was analyzed for content, sentiment, and thematic relevance using agenda-setting and uses and gratifications frameworks.

Results: Four dominant themes emerged: (1) patient experiences/testimonials, (2) drug pricing/insurance concerns, (3) pharmaceutical news, and (4) advertising commentary. The content shows a range of potentially agenda-setting topics, including pharmaceutical companies, their DTC drugs, patient experiences, costs, insurance, and advertising commentary. The uses and gratifications we found included information, entertainment, economic benefits, and social benefits, with information being the most prevalent.

Conclusions: Our study provides an important glimpse into what is being discussed on Twitter by, with, and about pharmaceutical drugs that spend billions on advertising each year [1]. Although more tweets were neutral than positive or negative, the overall sentiment of the top terms was negative. The predominance of negative sentiment in our findings suggests that many social media users express apprehensions or criticisms about these drugs, highlighting the need for improved communication, transparency, and engagement strategies. Further implications are also discussed surrounding public policy and industry practice. Suggested future research directions include NodeXL network analysis and qualitative inquiry into user motivations.

(Online J Public Health Inform 2026;18:e85641) doi:10.2196/85641

KEYWORDS

social media listening; direct-to-consumer pharmaceutical advertising; agenda setting; uses and gratifications; patient engagement

Introduction

Background

Around 60% of the world's population use social media [1], and surveys have shown that up to 26% of Americans use social media for health information [2]. In the United States, each year, pharmaceutical companies spend billions on direct-to-consumer (DTC) pharmaceutical drug advertising on various channels, with TV remaining the dominant medium by far [3]. DTC user-generated content on social media channels provides an interesting and important environment to study the conversations surrounding DTC drugs with high advertising budgets. DTC advertising, on any channel, has the potential to spark conversations among consumers. Those conversations on social media extend the word-of-mouth impact to a larger audience and provide a greater ability to monitor the conversation [4]. Recent publications express a continued need to better understand reactions and conversations about these drugs [5], given their large advertising budgets and the Food and Drug Administration (FDA)'s ongoing consideration of possible restrictions [6].

XSL•FO
RenderX

This study begins to illuminate the intersection between individuals' desire to obtain health information online, strategic communication firms' quest to promote clients' brands through creativity and engagement, and the FDA's mandate to oversee DTC drug advertising and protect public health. DTC drug conversations on social media have begun receiving more research attention, but one gap that remains is the use of social listening tools to better understand the content, sentiment, and trends surrounding DTC drugs on social media, beginning with those brands that spent the most from 2020 to 2022. This research contributes to scholarship in several ways, including agenda-setting research (what agendas might social media conversations set?). The managerial implications of what consumers are discussing online and how both advertising and public relations practitioners may need to address consumer concerns are also explored. Our research may help DTC advertisers understand the impact their creative output has, as expressed through consumers' social media conversations, uses, and gratifications. Policymakers and pharmaceutical companies may also benefit from identifying the agenda-setting themes of social media conversations. This research aims to advance our understanding of where and how individuals are exposed to DTC drug information from a variety of sources in our fragmented and diversified media environment. We will first review some of what is already known in this area before describing our specific focus and social listening methods.

DTC Advertising and User-Generated Content on Social Media

While TV still accounts for the vast majority of ad spending, the user-generated content and conversations surrounding DTC advertisements often occur online, and social listening facilitates our understanding of these massive datasets [7]. This study will help fill the gap in understanding what that online conversation consists of for the top spenders and will begin exploring the application of agenda-setting and uses and gratifications theories. As is often the case when an area is first being explored, the publications touch on a wide range of topics but do not include a broad content-based social listening study like ours. Papers that have been explored in this general area are summarized in Multimedia Appendix 1. After the inception of social media, research has investigated DTC on social media [8-11] to better understand broadly how pharmaceuticals were being promoted by their companies and how influencers may promote drugs [9], but little research has examined conversations and influencers used across drug categories. Several papers call for more research in this area [11,12].

Our study provides an important extension to this body of research by examining a broader sample of user-generated content, not just contents posted on company accounts or focused on only narrowly defined, albeit important, health concerns, such as obesity [13].

FDA Regulations of Social Media DTC Ads

Because our study helps document the content of social media DTC drug conversations, we briefly review the guidance documents to provide the reader with a general understanding of the broader FDA regulatory and historical context. We summarize key points of the FDA guidance documents and urge interested readers to read the documents for key details related to agency expectations. Especially for readers who may not be familiar with DTC advertising regulation, this provides needed context for our study of Twitter (subsequently rebranded X) conversations of high-budget DTC drugs and offers guidance for future research. FDA regulation in this area is ever evolving, and the social media chapter is still being documented.

The FDA wrote several guidance documents in the 90s, including those related to broadcast and print DTC ads. The essence of those was that the ads must not be false or misleading, should present a fair balance of benefits and risks, should include the most important risk information, as well as the product indication and limitations, in consumer-friendly language, and should provide access to more detailed information often via a website [14]. The FDA also made provisions for "reminder" (also known as help-seeking) ads that did not need to include risk information when the benefits were not included alongside the brand name.

One of the most obvious additions to the FDA guidelines to address social media messaging [15] was the handling of the character space limitations present on Twitter, renamed X in July 2023, and other social media platforms. Twitter, prior to 2017, was limited to 140 characters, but that limit doubled in 2017. In 2023, Twitter began enabling subscribers to tweet up to 4000 characters. Throughout the remainder of this paper, we will use "Twitter" and "tweet" since the vast majority of our data were collected when the platform used those terms. The FDA acknowledges that the character limit is challenging, given the amount of information needed for DTC advertising, and does allow a reference, often a link, to access full risk information, though risk information must still be balanced, compared to benefits, and as prominently displayed [15].

Due to the fluid nature of online promotion, the FDA also released a draft guidance to help communication agencies handle "postmarketing" materials, such as dynamic, real-time promotional activities online, including blogs or social networking sites [16]. The FDA distinguishes between 3 main types of content—sites directly under the control of the firm (pharmaceutical company or their communication firm), third-party sites, and content generated on the firm's behalf. In each of these instances, if the firm has even partial control (influence over content), then the firm is responsible for submitting to the FDA under the "postmarketing" submission requirements [16].

The third FDA draft guidance discusses third-party misinformation that may be posted online, and the FDA considers it voluntary whether the pharmaceutical company or its communication firm wants to correct the misinformation [17]. If the firm chooses to respond, the draft guidance provides more detail about how the response should be truthful and nonmisleading.

Theoretical Frameworks

Overview

Two theoretical frameworks were utilized to guide our research, analysis, results, and future research suggestions. We chose these 2 theories to best understand the varied roles of individuals

XSL•FO
RenderX

on social media: contributor to media content (agenda-setting) and consumer of content (uses and gratifications). Agenda-setting theory is useful for identifying which DTC drug topics are being discussed most frequently as these topics may be likely to impact engaged audiences, influence what they think about, and inform what industry, policymakers, and communicators may need to consider. Uses and gratifications theory has a natural fit with user-generated social content and provides a framework for understanding the content-related themes.

Agenda-Setting Theory

Agenda-setting theory has a history spanning over 50 years and encompasses multiple areas of mass communication research, including news media [18,19], social media [18], advertising [20], and health communication [21]. Agenda setting began as a way to describe how the news media establishes an agenda related to political campaigns and what the public considers important [22]. Decades of research have documented the strong relationship between the topics covered by the media and the issues people think are important [23]. Agenda setting provides a framework for understanding how social media may influence what individuals consider significant in relation to DTC drug advertising.

Although agenda-setting theory began with traditional news media, it has been applied to social media health promotion environments as well, specifically Twitter [4,21,24,25]. Yun et al [4] applied agenda setting to conversations on Twitter and examined key influencers to give recommendations for health campaigns based on their network analysis. This crucial area explores our understanding of which topics related to DTC drugs are being discussed and how they may influence individuals, the public, the media, organizations, and policy agendas.

Uses and Gratifications Theory

Uses and gratifications theory dates back to the 1940s, and the exploration aimed at better understanding why individuals choose certain media and what purpose it served them. Nearly 80 years and hundreds of publications later, Dolan et al [26] were among the first to suggest uses and gratifications theory's applicability to social media settings. They identified 4 main groups of gratifying content for social media users based on a review of published studies: information, entertainment, remunerative (economic) benefits, and relational (social) benefits. These differ somewhat in important ways from the earlier typology that included diversion, personal relationships, personal identity, and surveillance (information) [27]. Uses and gratifications theory has been successfully applied to social media in a variety of studies, including associations with problematic internet use or addiction [28], impact on social relationships and psychological well-being [29], and Facebook usage [30]. However, little to no research has specifically applied uses and gratifications theory to conversation involving DTC-advertised drugs or the pharmaceutical industry.

Broadly speaking, individuals may obtain information about DTC drugs directly from advertisements, traditional media channels, or social media. This study aims to help fill a gap by identifying topics that may both provide individuals with uses and gratifications and set agenda items for the public to consider. The review of theory and literature led us to formulate two key research questions:

RQ1: What user-generated health-related topics discussed on social media mention DTC pharmaceutical drugs with the biggest DTC ad budgets? Which, if any, could be agenda-setting to the broader audience?

RQ2: What are the possible uses and gratifications that individuals engage in related to DTC drugs on Twitter? Additionally, are the mentions positive, negative, or neutral based on sentiment analysis?

Methods

Our research used the social media analytics tool Brandwatch, which tracks online conversations across various social media channels. Social media "listening" has emerged as an important analytical method to investigate online social media content [31-33].

Ethical Considerations

The Texas Christian University Institutional Review Board reviewed the research protocol and found the study to be exempt for human subjects research (2023-202). The authors adhered to personal, professional, and research ethics in conducting this research and preparing the manuscript, and they did not discern any conflict with the journal's stated ethics. No informed consent or compensation was required as the review was exempt from human subjects review.

Sampling

Given that very little research has applied social media listening techniques to DTC brands, we used purposive sampling and included the top DTC advertising spenders for 2020 to 2022, according to Kantar and Vivvix (listed in alphabetical order): Botox, Cibinqo, Dupixent, Eliquis, Entyvio, Humira, Jardiance, Nurtec ODT, Opdivo (+Yervoy), Otezla, Ozempic, Rexulti, Rinvoq, Rybelsus, Skyrizi, Tremfya, Trulicity, Verzenio, and Xeljanz [3,34-36]. Our rationale is that these brands may be most likely to be talked about on social media, given the preponderance of ads on traditional media. Two brands that spend significantly on digital formats, even with smaller overall budgets, were added to the sample to provide a broader, more diverse perspective—Orgovyx and Ponvory [3]. Multimedia Appendix 2 includes the study's drugs, their primary condition treated, and their parent companies.

Study Design and Query Construction

We collected historical and archived tweets from August 1, 2021, to August 1, 2023, related to a sample of top-spending DTC drug brands from 2020 to 2022 (Cibingo, Dupixent, Eliquis, Entyvio, Humira, Jardiance, Nurtec ODT, Yervoy, Otezla, Rexulti, Rinvoq, Rybelsus, Skyrizi, Tremfya, Trulicity, Verzenio, Xeljanz, Orgovyx, Ponvory, Botox, and Ozempic). We used these strategic keywords (drug brand names) as a query. We filtered the tweets to show only those in English and posted by Twitter accounts in the United States. Our first query gathered 486k total mentions and 247k unique authors.

XSL•FO
RenderX

However, when analyzing the results, we found that most of these mentions came from only the keywords “Ozempic” and “Botox,” and these tweets focused on mentions not relevant to the drugs themselves or our research questions, such as popular culture or unrelated celebrity commentary, adversely affecting the remaining keywords’ results by masking the DTC pharmaceutical-oriented conversation. We removed Botox and Ozempic to facilitate a clearer analysis by eliminating noise caused by discussions that did not pertain to the research focus. It is important to acknowledge potential limitations in discerning the context of conversations related to these drugs, but the abundance of popular culture references unrelated to the study’s focus overshadowed the relevant mentions from other brands under examination and skewed the results. As a result, the keywords “Ozempic” and “Botox” were no longer included in the final query and sample.

After analyzing the query results, we found that spam tweets, deals and promotions, and automated bot tweets were also included in these mentions. Therefore, we refined the query and used Boolean operators (OR and AND NOT) to combine and exclude keywords and spam tweets that brought irrelevant results to the search (see Multimedia Appendix 3 for additional details on query construction) [37]. The new search results gathered a more appropriate number of mentions, 44.7k, for the second query during the selected time period. The sampling method involved capturing tweets over a specified period using relevant keywords to ensure a representative dataset. Sentiment analysis was conducted using Brandwatch’s natural language processing algorithms, which were further validated by a manual review to correct any discrepancies. Author analysis included examining tweet authors’ profiles to understand the contributors’ demographics and influence.

Brandwatch generated helpful data reports and organized them based on the search query topics into 3 main categories: measure, conversation, and people. The measure category included graphs that showcased the volume, content sources, top sites, sentiment, and emotion of all tweets. Additionally, Brandwatch generated a topics wheel, word cloud, and list of trending topics and top hashtags for the conversation category. Finally, the “People” section presented the general demographics of the Twitter users, such as gender, interests, profession, and influencers for the tweets in our data.

Statistical Analysis

Topics Analysis

The topics wheel collected the most frequently mentioned topics in an interactive wheel graph. Each topic could be clicked to view a list of all the tweets containing that specific keyword. Similarly, the interactive word cloud displayed all the words, phrases, and entities most frequently found within the tweets. In a word cloud, the font size of a word is determined by how often the word is mentioned across tweets. The more frequently a word is used, the larger it appears in the word cloud. Like the topics wheel, clicking on each keyword showed the total number of mentions and the tweets for each mention. Each tweet included its publishing date, reach, and a link to the Twitter post and account profile.

We further analyzed the data using visualization methods (charts or other graphic tools) to discover patterns across mentions by applying thematic analysis techniques to the qualitative data captured through Brandwatch. Initially, Brandwatch’s artificial intelligence feature, Iris, generated preliminary categories. We then meticulously reviewed, reorganized, and refined these categories to ensure the accurate classification of sentiments and themes. This process followed the inductive, emergent approach recommended by Strauss and Corbin [38], involving open coding to identify initial themes and axial and selective coding to refine these themes into core categories. The topics for coding were selected based on their frequency and relevance to the discussion on Twitter about DTC pharmaceutical drugs. Relevant tweets included any mention that helped determine the type of content discussed on Twitter regarding DTC pharmaceutical drugs.

Custom Classifiers

We then used custom classifiers, Brandwatch’s machine-learning tool that uses the Brightview algorithm, to train custom classification models by adding mentions from our query results into different categories. The authors used a process similar to content analysis to agree upon how exemplar data should be classified. Once trained, the algorithm automatically assigns categories to each mention. The four main topics identified in the query for all other DTC drugs are (1) patient experiences, testimonials, or opinions; (2) high drug costs, insurance, or medical coverage; (3) pharmaceutical company or drug news; and (4) advertising discussions.

Sentiment Analysis

To determine attitudes, opinions, and emotions expressed by Twitter users, sentiment was analyzed using the Brandwatch sentiment score of the 50 most frequently used keywords, also shown in the word cloud (see Multimedia Appendix 4). Brandwatch determines the sentiment metric by weighing the number of positive and negative tweets for the keywords. A higher net sentiment indicates more positive than negative mentions, and a lower net sentiment indicates a higher ratio of negative to positive mentions.

Brandwatch identifies words and phrases that indicate positive or negative sentiment while setting rules that consider the effect context might have on the content’s tone. Because Brandwatch’s automated sentiment analysis process may miss the nuances of human communication (eg, sarcasm), we checked the positive, negative, and neutral mentions to verify that they were properly categorized.

Sentiment analysis is valuable for understanding public attitudes, opinions, and emotions on social media [39]. Analyzing DTC pharmaceutical advertising sentiment can provide crucial insights into how consumers perceive drug brands, their effectiveness, and associated risks. Research has shown that consumer emotions significantly influence decision-making, including health-related choices [40]. By examining sentiment in Twitter discussions, we can identify potential concerns, such as skepticism about a drug’s efficacy, fears of side effects, or distrust of pharmaceutical companies.

XSL•FO
RenderX

Brandwatch's sentiment analysis uses advanced natural language processing techniques through machine learning and linguistic models to interpret social media conversations in context [40]. These models are designed to recognize slang, misspellings, and regional dialects, making them more effective at accurately classifying positive, negative, and neutral emotions. However, given the potential limitations of automated sentiment classification, manual verification is crucial for refining accuracy and reducing misclassification errors.

Sample Description of Authors' Posting

Understanding the demographic information of tweet authors is helpful for providing context to online conversations. Brandwatch determines the gender distribution of tweet authors by analyzing user profile details and post content. However, the exact methods used for gender inference remain proprietary and are not publicly disclosed.

The data provided by Brandwatch about the authors who posted the Twitter content help provide context about the content we described. Slightly more than half of the posts were categorized as men (Multimedia Appendix 5). The most common professions associated with these Twitter accounts included health practitioners, artists, executives, and teachers/lecturers. The authors' top interests were family and parenting, politics, books, and beauty/health and fitness. The most mentioned influencers and authors included @BethWaldron (294 mentions), @pharmacychecker (243 mentions), @TheAsthmaCures (163 mentions), @newsfilterio (144 mentions), and @CrweWorld (102 mentions).

Results

Making Sense of the Data

We will first discuss the results of RQ1—what are the user-generated health topics that may be agenda setting—followed by RQ2—how might these topics suggest individuals' uses and gratifications, including sentiment analysis. Between August 1, 2021, and August 1, 2023, there were a total of 44.7k mentions and 26.8k unique authors on Twitter for the search query related to the top-spending DTC drugs. Although mentions included earned, shared, and owned media about the topics of interest, the majority of mentions were shared and earned. The word cloud in Multimedia Appendix 4 included the top 50 most frequently used terms in the Twitter mentions for this time period. In addition to the size of the word, the shading of the word cloud indicates the volume of tweets associated with each keyword. The darkest shade of blue represents the words with the highest volume, while the lightest shade of blue represents those with the lowest volume.

We analyzed the most common words mentioned within the Twitter results through the word cloud derived from Brandwatch Insights. The top 10 most frequently used terms (Multimedia Appendix 6), contributing to a large majority of all mentions, are "drug" (5796 tweets), "month" (4276 tweets), "years" (3501 tweets), "cost" (3474 tweets), "dose" (3327 tweets), "COVID" (3092 tweets), "treatment" (2911 tweets), "eczema" (2845 tweets), "patients" (2845 tweets), and "vaccine" (2764 tweets). It is important to note that the number of mentions for the most frequently used words is not cumulative. For instance, if a single post includes the words "drug," "month," and "cost," it will be counted as 3 separate mentions—1 for each keyword—despite originating from a single tweet.

Many keywords included mentions that covered more than one of the identified categories. Using Brandwatch's OpenAI tool, Iris, and a close examination of the keywords in the word cloud, we found four main topics from the mentions gathered: (1) patient experiences, testimonials, or opinions; (2) high drug costs, insurance, or medical coverage; (3) pharmaceutical company or drug news; and (4) advertising discussions. Top keywords for each topic are listed in Textbox 1.

Textbox 1. Keywords categorized by topics.

1. Patient experiences, testimonials, or opinions
   - side effects, years, patients, eczema got worse, blood, medication, taking, week, people
2. High drug costs, insurance, or medical cover
   - COVID, dermatologist, month for dupixent, vaccine dose, told, years, paying, pocket, costs
3. Pharmaceutical company or drug news
   - drug, cost, US, United States, cancer, dollars, prescription, biosimilar, treatment, lower
4. Advertising discussions
   - commercial, song, ad, jingle, diabetes, woman, TV, people, Jardiance, commercials, big

RQ1: Agenda-Setting Twitter DTC Drug Themes

Patient Experiences, Testimonials, or Opinions

The results from the keywords "years," "patients," "week," "side effects," "migraine," "eczema got worse," "told," "life," "blood," "fed," "medication," and "control" also include a wide range of patient experiences and testimonials. Additionally, users discuss the side effects of various medications, such as Nurtec ODT, Humira, Jardiance, Rexulti, Xeljanz, Dupixent, Verzenio, and Eliquis. Concerns are expressed about the side effects of biologic medications, particularly Humira and Dupixent. Multiple users shared their experiences and side effects of taking drugs like Ozempic, detailing symptoms such as pale skin, lack of energy, constant body aches, and nausea. Notably, despite removing Ozempic from the sample due to its

XSL•FO
RenderX

prevalence in popular culture references not relevant to the study, the term was still mentioned alongside other drug names. These mentions differed significantly as they directly reflected the experiences of patients, offering valuable insights into comparative analyses with other brands previously utilized by patients.

However, positive experiences with medications are also highlighted, with some tweets emphasizing the transformative effects of certain drugs, such as Dupixent improving the lives of individuals with eczema and Nurtec ODT providing relief for people experiencing migraine, overall improving their quality of life. For example, user @E_lisa_Beth tweeted, “I have had eczema since I was 2. I’ve had it just about everywhere. It got to the point my hands would split open. I couldn’t wear any type of sandals, b.c my feet were so bad. Been on dupixent for almost 4 years now. It is the miracle medication I’ve been waiting for my entire life.”

High Drug Costs, Insurance, or Medical Coverage

The keywords and phrases “cost,” “dermatologist,” “vaccine dose,” “pocket,” “month for Dupixent,” “years,” “paying,” “told,” and “COVID” were some of the most frequently mentioned words in conversations on Twitter that highlight a range of issues related to high drug prices, the cost of medications, and insurance dissatisfactions. Users compared drug prices, emphasizing the discrepancy in the cost of Humira between the United States and other countries, including Italy, Germany, and France. They also criticized pharmaceutical companies for what is perceived as price gouging, particularly in the case of top-selling drugs, such as Eliquis, Humira, Trulicity, and Revlimid, which, according to users, are significantly more expensive in the United States than in other countries (eg, @ChrisMurphyCT posted on Twitter “Cost of Humira: US: $3216 Italy: $526 Germany: $420 France: $248 When are we going to realize that our drug pricing policies just subsidize cheap drugs for the rest of the world? It’s a classic free rider problem.”).

A common topic to tweet about is the financial burden placed on patients because of high medication costs and the need for more affordable options. Furthermore, there is criticism of US drug pricing policies, with users suggesting that these policies result in subsidizing cheaper drugs in other countries. These comparisons highlight the substantial differences in drug pricing and raise questions about why drugs are more expensive in the United States compared to other nations. Some tweets mention the role of government, policymakers, and legislation in addressing the issue of high drug prices and negotiating prices with pharmaceutical companies to make medications more affordable for patients (eg, @Terryfor88 tweeted, “@amyklobuchar Could you look into the cost of Eliquis, a lifesaving necessary drug. It is very expensive even with Medicare and a supplement. There is no generic so drug companies need to lower the cost. Thank you.”).

Some individuals expressed their frustration with out-of-pocket expenses and the impact on their ability to afford necessary treatments, even with insurance coverage. Insurance and medication coverage concerns are also prevalent in Twitter conversations with the keywords “U.S.,” “United States,” “pocket,” and “insurance.” Users discussed issues with insurance coverage for medications like Eliquis and the need to switch to alternative drugs due to insurance policies. There were mentions of insurance companies denying coverage or requiring prior authorization for specific medications, leading to dissatisfaction among individuals regarding insurance plans and their handling of drugs like Entyvio, Humira, Mounjaro, Victoza, and Trulicity. For example, @LiliVicCreation tweeted, “I’m absolutely gutted about this Dupixent situation. My insurance is already denying my refill requests and I have no more medicine. I can’t believe these companies can just decide to stop covering medications. I hate private insurance companies.”

Pharmaceutical Company, Drug News, and Advertising Discussions

Because these 3 topics were somewhat intertwined, we present them in one combined section. Some Twitter discussions, including keywords such as “U.S.,” “United States,” “dollars,” “commercial,” “best-selling,” “biosimilar,” “Jardiance commercials,” and “treatment,” concerned pharmaceutical company practices and competition. Criticism was directed at pharmaceutical companies, particularly AbbVie, for increasing drug prices over time and delaying competition in the market. Users also expressed concerns about the implications of competition among drug companies, particularly blockbuster drugs such as Humira and Dupixent.

Moreover, some tweets expressed mixed feelings about the role of celebrity endorsements in pharmaceutical marketing, such as Lady Gaga endorsing Pfizer’s migraine drug Nurtec ODT, with some people asking Lady Gaga to “stop pushing pharmaceuticals on the public and her fans.” Additionally, criticism of the portrayal of individuals with overweight conditions was found in certain pharmaceutical commercials, leading to discussions about insulin costs and misconceptions about diabetes. For example, a user tweeted, "I’m sure they mean well but everyone in the new Jardiance commercial is obese. This fuels GOP insulin costs talking points. Type 1 diabetics are insulin dependent & it has nothing to do with weight. It runs in my family. My slim brother was dx with type 1 diabetes @ age 28.”

On the other hand, amusement and mockery of what users claimed to be “the cheesy and irritating” nature of pharmaceutical commercials made them “cringe.” Mentions revolve around the annoyance, particularly focusing on drugs like Jardiance’s and Skyrizi’s use of jingles and repetitive messaging. They also discussed how commercials often highlight countless side effects, pricing issues, and the overall saturation of drug advertisements in the media.

Users express mixed feelings about the Skyrizi ad, with some finding it amusing while others feel overwhelmed by its frequency. The commercial’s focus on various medical conditions and its musical approach has sparked discussions and reactions among viewers. A user wrote, “I still hate that Jardiance commercial. The little pill I hope goes straight to hell.” There are also discussions about the pharmaceutical industry’s advertising practices, as seen in commercials for drugs such as Jardiance or Skyrizi. The ads’ focus on specific medical conditions, side effects, and pricing issues has sparked

XSL•FO
RenderX

debates about the effectiveness and ethics of DTC marketing in the health care sector.

The keywords and phrases “FDA,” “years,” “COVID,” “patients,” “effects,” “prescription,” “biosimilar,” “control,” and “cancer” are some that highlight various regulatory aspects of FDA approvals and drug treatment news, including those for different types of cancer and chronic diseases. Additionally, multiple tweets reported FDA updates regarding various medications, including label updates, approvals for treating specific conditions, diseases, or new indications, and warnings about compounded medicines and their use as substitutes for prescription drugs. Some tweets from news accounts announced the FDA’s approval of medications such as Rexulti and Dupixent for treating symptoms associated with dementia, specifically agitation and eosinophilic esophagitis, respectively. Other tweets announced the approval of drugs such as Rinvoq and Rexulti for treating particular diseases or conditions, such as Crohn disease, eosinophilic esophagitis, agitation in individuals with Alzheimer disease, and psoriatic arthritis (@LEAD_coalition tweeted, “FDA Approves First Drug to Treat Agitation Symptoms Associated with #Dementia due to #Alzheimer Disease https://www.fda.gov/news-events/press-announcements/fda-approves-first-drug-treat-agitation-symptoms-associated-dementia-due-alzheimers-disease by @US_FDA.”).

RQ2: Topic Application to Uses and Gratifications

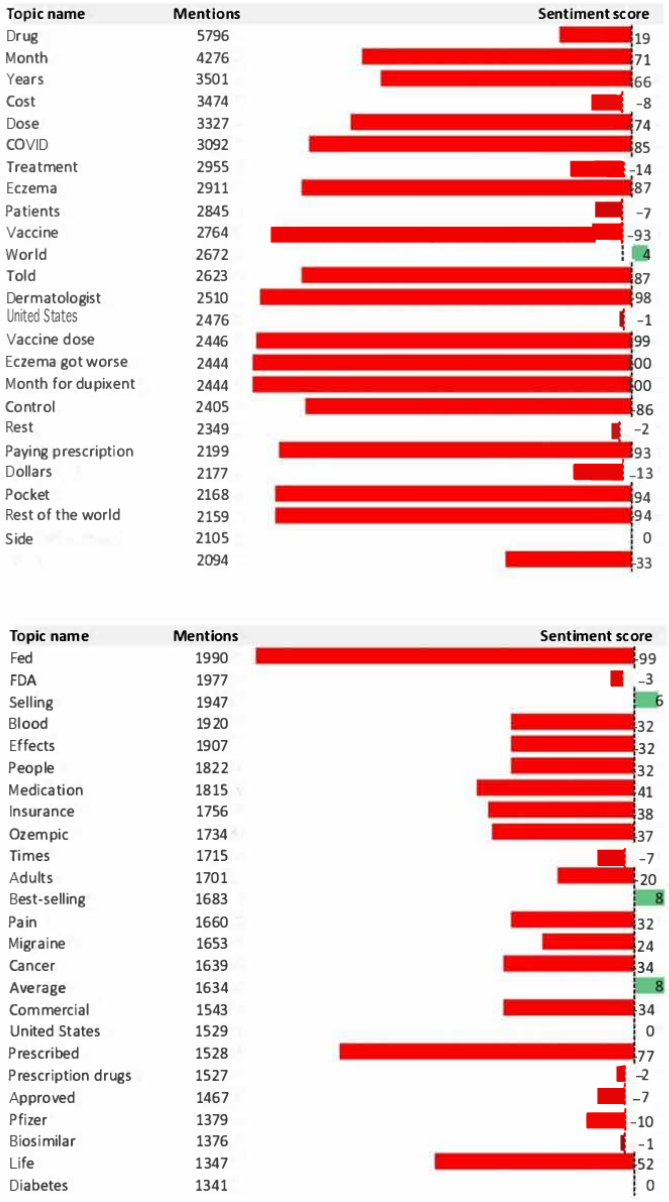
Sentiment Analysis

Examining how the topical themes indicate individuals’ uses and gratifications provides an additional layer of analysis and application. The 4 uses and gratifications we used came from the social media perspective, including information, entertainment, economic benefits, and social benefits [28], due to their applicability to the unique social media environment. We felt it was important to consider the sentiment analysis here to capture the positive or negative nature of these posts, as it helps capture the context of how the giver or receiver is feeling and the underlying tone of the posts.

The overall net sentiment score for the top 50 keywords is negative, with 43 keywords having an average sentiment score of –47.7, 3 keywords with a score of 0, and 4 keywords with an average sentiment score of 6.5. Figure 1 shows each keyword’s total number of mentions and sentiment score.

XSL•FO
RenderX

Figure 1. Query keyword mentions and sentiment score. FDA: Food and Drug Administration.



Information

Some express a neutral sentiment, focusing on competition among pharmaceutical companies, including discussions about alternative drug options or biosimilars [41]. However, most neutral mentions encompass a wide range of earned media shared on Twitter, highlighting the achievements of pharmaceutical companies, drug advancements, or research, without indicating a favorable or unfavorable stance toward the subject matter. For instance, Fierce Pharma tweeted,

XSL•FO
RenderX

"Boehringer Ingelheim posts more Jardiance growth as potential US kidney disease approval nears," while linking their news article.

The effectiveness of various medications in treating conditions such as rheumatoid arthritis and colitis was another discussion reflecting some patients' negative experiences. We listed these under information because individuals were often sharing their own experiences to help others who might benefit from knowing. Positive tweets mentioned users' positive experiences and results with medications, such as Skyrizi, Humira, Rinvoq, and Trulicity. A user responded to a tweet that asked for suggestions for autoimmune illness-related eczema by describing how they were "grateful every day that I take Rinvoq for my autoimmune illness. Makes a world of difference. Five earlier meds were useless. Keep trying." Similarly, another user replied, "I have been on Dupixent for eczema, and I also have RA so I very much feel you. Been on the Dupixent for just over a year, and it cleared 99% of my eczema which was mostly on my hands within a month or two. Absolute game changer for me and the only thing that's ever worked."

Entertainment

Tone was important for entertainment as some posts were sarcastic or made fun of the ads or creative elements within them. It might not be that the individuals found the ad entertaining but rather that they were attempting to entertain others with their commentary. Example tweets are included in Textbox 2.

Textbox 2. Jardiance posts.

The commercial for *Jardiance* with the lady dancing and singing is so stupid. How does one have a "touch" of diabetes?
Ok I hate Pharma and I HATE Pharma advertising but real talk:
that *Jardiance* commercial kinda slaps "I'm lowering myyyy alc!!! Here's Hazel enjoying the commercial. User added a picture of their cat looking at the TV with the commercial in it.

Economic Benefits or Concerns

Many of the negative mentions gathered in this analysis primarily focused on different aspects of health care and medication concerns, aligning with economic benefits—or in this case, concerns. Notable themes include reports of negative side effects following COVID vaccinations, particularly in the context of eczema worsening, which lead to costly prescriptions like Dupixent for symptom management (Figure 2). Medication prices and their impact on affordability were a prominent concern, with many individuals expressing frustration and apprehension about the high costs associated with drugs such as Humira, Eliquis, Jardiance, and Trulicity, sometimes resulting in discontinuation (ie, ceasing medication treatment) or financial hardships. Insurance coverage issues compounded the financial burden, particularly when obtaining specific drugs such as Humira and Entyvio. Figure 3 shows a post stating Humira is "one of the best-selling…" but costs most in the United States than the rest of the world.

Figure 2. Dupixent screenshot.

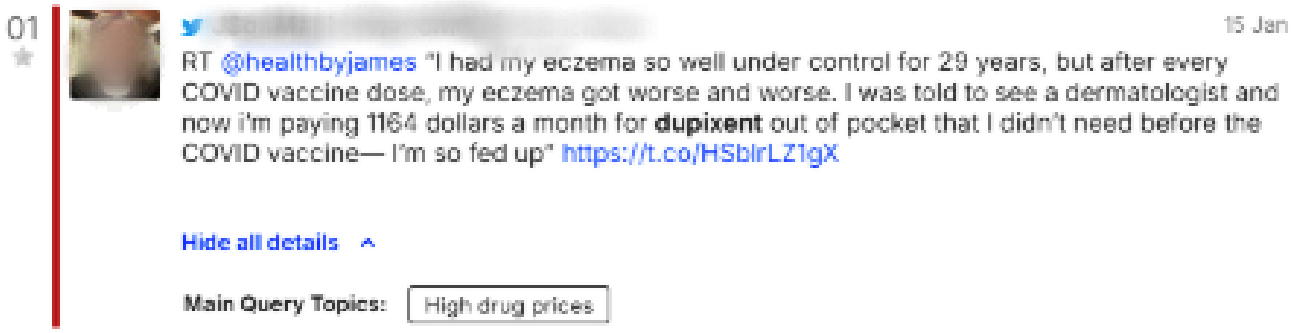
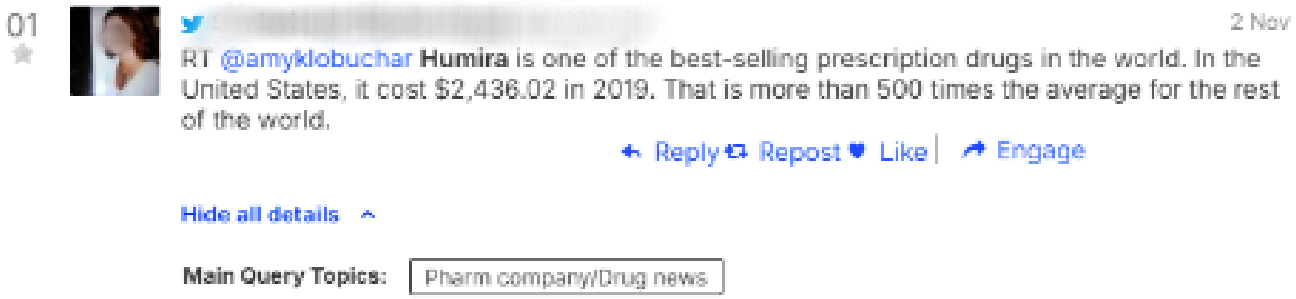


Figure 3. Humira screenshot.



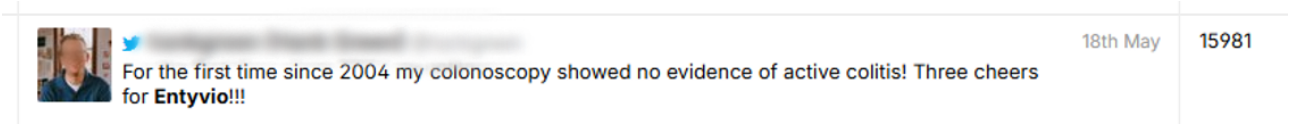
Tweets also highlight users' positive results from drug trials, such as Dupixent and Nurtec ODT, in treating chronic diseases, including chronic obstructive pulmonary disease, migraines, asthma, eczema, and more. These could honestly fit under information or social benefits, depending on the context of various cases.

XSL•FO
RenderX

Social Benefits

Social benefits were often related to posts about how the drugs impacted their lives in a positive way or how grateful they were for the drug treatment, as shown in Figure 4.

Figure 4. Entyvio screenshot.



While these were the only topics clearly tied to social and relational benefits, we know that there may be unspoken social benefits inherent in social media interactions; however, the tweets themselves did not display that content beyond these instances for our data and topic area.

Discussion

Our study provides an important glimpse into what is being discussed on Twitter by, with, and about pharmaceutical drugs, which spend billions on advertising each year [3]. Although more tweets were neutral than positive or negative, the overall sentiment of the top terms was negative. The predominance of negative sentiment in our findings suggests that many social media users express apprehensions or criticisms about these drugs, highlighting the need for improved communication, transparency, and engagement strategies. These data are particularly valuable in pharmaceutical advertising, as they help identify consumer concerns, skepticism, and potential misinformation [38].

The content highlights a range of potentially agenda-setting topics, including pharmaceutical companies, their DTC drugs, patient experiences, costs, insurance, and advertising commentary. Twitter is serving as a forum for pharmaceutical companies, their constituents, and social media health influencers to discuss topics related to top-spending DTC drugs. To make the 44.7k mention number less arbitrary, our data indicate an average of 61 top-spending DTC drug mentions per day. We have certain hot topics and icons like Lady Gaga garner far more, but Twitter seems to be a popular place to connect with others about pharmaceutical concerns and thoughts.

We found both agenda-setting and uses and gratifications theories helpful for contextualizing our data. On Twitter, since anyone can post, the concept of agenda-setting broadens to include a variety of experts and nonexperts. The result is often a collection of opinions, perceptions, experiences, facts, and fictions. Key agendas emerge in the topical themes we identified regarding pharmaceutical companies, insurance, drug costs, and patient experiences. However, intent is not always clear; do people post with the intent of influencing a broader agenda? Some may want to share experiences for the cathartic nature of it, while others aim to influence the system or help others on their path. This need to connect has been noted in previous research [41] about online message boards, which were the precursors to social media.

This study demonstrates how Brandwatch can provide a high-level overview of the DTC drug conversation on social media. None of the topics are a surprise, but they empirically confirm what we may anecdotally believe is on the minds of consumers and patients. Professional DTC drug communicators need to be in tune with their audiences so that consumers do not perceive their messages as oblivious, out of touch, and only focused on their brand. We observe this disconnect between communication strategy (eg, singing and dancing) and patients' headspace in the comments about "cheesy and irritating" nature of DTC ads that has also been noted in academic research [42,43]. DTC ads have a very challenging communication situation in balancing FDA guidance requirements, conveying technical medical terms in consumer-friendly ways, executing the branding strategy, satisfying client wishes, and connecting with an often-diverse audience. Tools such as Brandwatch and social media listening are key to staying relevant and up to date.

Uses and gratifications were not equally represented but were dominated by information, economic benefits, and entertainment. It was not evident how social benefits were sought or obtained. These findings are helpful for both academic researchers to understand the applications of this longstanding theory, as well as for policymakers to be aware of their constituents' concerns. Some advertising practitioners, especially those with DTC drug accounts, may also benefit if they do not have access to tools such as Brandwatch. The data presented in this study help them to understand what topics are most often discussed for high-budget brands, as well as the uses and gratifications that social media users predominantly obtain.

This study also has implications for policymakers who oversee the health insurance, medical, and pharmaceutical industries. We found several themes related to regulation and pricing in these industries that policymakers and regulators would benefit from understanding better. DTC drug advertising is at this time only allowed in the United States and New Zealand and has garnered criticism since its start in the 1990s for increasing prices, among other things [44,45]. It is interesting that comments attributed high drug costs in the United States to subsidizing drugs in other countries as opposed to the cost of DTC contributing to what consumers pay.

Important next steps in this area of research may include a network analysis to examine the connections and interactions between various agents (eg, pharmaceutical companies, social media influencers, individuals, etc). This would reveal how the information flows through networks since our investigation focused solely on the content of the information. Another extension may be to include a wider variety of media budget beyond the top spenders among DTC drugs and see how volume and type of conversations may differ. A final extension may be to explore the qualitative and quantitative data to understand the why behind the what; since social media listening is similar to content analysis and cannot know why something was written. We encourage researchers to continue the application of

XSL•FO
RenderX

agenda-setting and uses and gratification theories, especially in pursuit of cause-and-effect findings. Social media listening is most similar to content analysis and is not able to know what effect the content topics we documented are. Do the topics impact what consumers care about (agenda setting)? Do policymakers pursue legislation related to these key agendas? What is the motivation behind individuals' uses and gratifications? Do these conversations impact consumer choice related to health conditions or DTC drug behaviors (eg, medical professional conversations, prescription adherence, or interpersonal word-of-mouth referrals)?

These ideas for future research would help to fill gaps left by the limitations of this study. Since this study focused on top-spending DTC drugs, it is important to look at a range of different spending levels and draw a probability sample to get a representative view of all DTC social media drug postings. It is also important to keep in mind that drug brands with a high advertising budget typically have also recently been approved for medical use by the FDA, act in a highly competitive market, and/or are under high public scrutiny. This means that a high advertising budget alone does not necessarily explain characteristics of the social media discourse on Twitter, and these other factors may also contribute to this study's findings. Expanding the scope to include more platforms, beyond Twitter, would be another logical next step to compare differences and similarities. Furthermore, studies examining the content of messages are limited by not knowing what impact those messages have in terms of interaction, awareness, and understanding. These are key areas for future research. Future research may examine the connections between posts, topics, brands, and entity posting, perhaps using NodeXL or similar tools for conversation analysis. Qualitative investigations could explore the strategies or motivations of the individuals posting to delve even deeper.

A couple of other limitations the reader should consider include the sample description, historical events during the sample timeframe, and the difficulty in assessing sentiment. While estimates of tweet author descriptions promise useful insight into demographic trends, we must also recognize that this methodology may have inherent limitations. The categorization of keywords was not mutually exclusive, so the reader will notice that keywords could be shared between topic categories, resulting in an unavoidable overlap in naturally adjacent topics, such as patient experiences and drug costs.

Acknowledgments

The authors acknowledge colleagues' comments about this manuscript that contributed to its present form. The authors state that they did not use any artificial intelligence tools except those integrated into Brandwatch.

Funding

The authors declare no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research involving direct-to-consumer advertising and user-generated content on social media.
[DOCX File, 27 KB - ojphi_v18i1e85641_app1.docx]

Multimedia Appendix 2
Dataset of direct-to-consumer prescription drugs.
[DOCX File, 18 KB - ojphi_v18i1e85641_app2.docx]

Multimedia Appendix 3
Details about the Brandwatch query.
[DOCX File, 19 KB - ojphi_v18i1e85641_app3.docx]

Multimedia Appendix 4
Top 50 Twitter keywords in query data.
[PNG File, 251 KB - ojphi_v18i1e85641_app4.png]

Multimedia Appendix 5
Description of authors.
[PDF File, 47 KB - ojphi_v18i1e85641_app5.pdf]

Multimedia Appendix 6

XSL•FO
RenderX

Top 10 Twitter keywords for query data.
[PNG File, 67 KB - ojphi_v18i1e85641_app6.png]

References

1. 4 pioneering social media campaigns in pharma this year. PharmaVoice. URL: https://www.pharmavoice.com/news/best-pharma-social-media-campaigns/636629/ [accessed 2023-07-14]
2. Aikin KJ, Sullivan HW, Berktold J, Stein KL, Hoverman VJ. Consumers' experience with and attitudes toward direct-to-consumer prescription drug promotion: a nationally representative survey. Health Mark Q 2021;38(1):1-11. [doi: 10.1080/07359683.2021.1947067] [Medline: 34238134]
3. The top 10 pharma drug ad spenders for 2022. Fierce Pharma. URL: https://www.fiercepharma.com/special-reports/top-10-pharma-drug-brand-ad-spenders-2022 [accessed 2023-07-11]
4. Yun GW, Morin D, Park S, et al. Social media and flu: media Twitter accounts as agenda setters. Int J Med Inform 2016 Jul;91:67-73. [doi: 10.1016/j.ijmedinf.2016.04.009] [Medline: 27185510]
5. Kravitz RL. Health care ramifications of pervasive direct-to-consumer prescription drug advertising. Annu Rev Public Health 2026 Apr;47(1):479-497. [doi: 10.1146/annurev-publhealth-081524-111735] [Medline: 41052312]
6. How the FDA's crackdown on pharma ads may play out legally. MM+M (Medical Marketing and Media). URL: https://www.mmm-online.com/news/how-the-fdas-crackdown-on-pharma-ads-may-play-out-legally/ [accessed 2026-06-02]
7. Liu X, Burns AC, Hou Y. An investigation of brand-related user-generated content on Twitter. J Advert 2017 Apr 3;46(2):236-247. [doi: 10.1080/00913367.2017.1297273]
8. Carpentier FRD. Considering the future of pharmaceutical promotions in social media comment on "Trouble spots in online direct-to-consumer prescription drug promotion: a content analysis of FDA warning letters". Int J Health Policy Manag 2016 Feb 9;5(4):283-285. [doi: 10.15171/ijhpm.2016.15] [Medline: 27239874]
9. DeAndrea DC, Vendemia MA. How affiliation disclosure and control over user-generated comments affects consumer health knowledge and behavior: a randomized controlled experiment of pharmaceutical direct-to-consumer advertising on social media. J Med Internet Res 2016 Jul 19;18(7):e189. [doi: 10.2196/jmir.5972] [Medline: 27435883]
10. Fogel J, Shraybman-Buynova J. Direct-to-consumer prescription medication advertisements on social media: the role of social factors. Health Mark Q 2023;40(2):190-205. [doi: 10.1080/07359683.2021.2011998] [Medline: 34866544]
11. Willis E, Delbaere M. Patient influencers: the next frontier in direct-to-consumer pharmaceutical marketing. J Med Internet Res 2022 Mar 1;24(3):e29422. [doi: 10.2196/29422] [Medline: 35230241]
12. Willis E, Friedel K, Heisten M, Pickett M, Bhowmick A. Communicating health literacy on prescription medications on social media: in-depth interviews with "patient influencers". J Med Internet Res 2023 Mar 13;25:e41867. [doi: 10.2196/41867] [Medline: 36912881]
13. Haggerty T, Sedney CL, Cowher A, Holland D, Davisson L, Dekeseredy P. Twitter and communicating stigma about medications to treat obesity. Health Commun 2023 Dec;38(14):3238-3242. [doi: 10.1080/10410236.2022.2144303] [Medline: 36373192]
14. Guidance for industry: consumer-directed broadcast advertisements. : U.S. Food and Drug Administration (FDA), U.S. Department of Health and Human Services; 1999 URL: https://www.fda.gov/media/75406/download [accessed 2026-05-29]
15. Draft guidance for industry: internet/social media platforms with character space limitations—presenting risk and benefit information for prescription drugs and medical devices. : U.S. Food and Drug Administration (FDA), U.S. Department of Health and Human Services; 2014 URL: https://www.fda.gov/media/88551/download [accessed 2026-05-29]
16. Draft guidance for industry: fulfilling regulatory requirements for postmarketing submissions of interactive promotional media for prescription human and animal drugs and biologics. : U.S. Food and Drug Administration (FDA), U.S. Department of Health and Human Services; 2014 URL: https://www.fda.gov/media/87685/download [accessed 2026-05-29]
17. Guidance for industry: internet/social media platforms: correcting independent third-party misinformation about prescription drugs and medical devices. : U.S. Food and Drug Administration (FDA), U.S. Department of Health and Human Services; 2014 URL: https://www.fda.gov/media/88545/download [accessed 2026-05-29]
18. Feezell JT. Agenda setting through social media: the importance of incidental news exposure and social filtering in the digital era. Polit Res Q 2018 Jun;71(2):482-494. [doi: 10.1177/1065912917744895]
19. McCombs M. A look at agenda-setting: past, present and future. J Stud 2005;6(4):543-557. [doi: 10.1080/14616700500250438]
20. Luo Y, Burley H, Moe A, Sui M. A meta-analysis of news media's public agenda-setting effects, 1972-2015. J Mass Commun Q 2019 Mar;96(1):150-172. [doi: 10.1177/1077699018804500]
21. Albalawi Y, Sixsmith J. Agenda setting for health promotion: exploring an adapted model for the social media era. JMIR Public Health Surveill 2015;1(2):e21. [doi: 10.2196/publichealth.5014] [Medline: 27227139]
22. McCombs ME, Shaw DL. The agenda-setting function of mass media. Public Opin Q 1972;36(2):176. [doi: 10.1086/267990]
23. McCombs M, Stroud NJ. Psychology of agenda-setting effects. Mapping the paths of information processing. Rev Commun Res 2014;2(1):68-93. [doi: 10.12840/issn.2255-4165.2014.02.01.003]
24. Dadich A, Khan A. Using Twitter to promote a youth mental health agenda. Health Promot Int 2021 Mar 12;36(1):235-249. [doi: 10.1093/heapro/daaa016] [Medline: 32380522]

XSL•FO
RenderX

25. Tahamtan I, Potnis D, Mohammadi E, Singh V, Miller LE. The mutual influence of the World Health Organization (WHO) and Twitter users during COVID-19: network agenda-setting analysis. J Med Internet Res 2022 Apr 26;24(4):e34321. [doi: 10.2196/34321] [Medline: 35275836]
26. Dolan R, Conduit J, Fahy J, Goodman S. Social media engagement behaviour: a uses and gratifications perspective. J Strat Market 2016;24(3-4):261-277. [doi: 10.1080/0965254X.2015.1095222]
27. McQuail D. McQuail's Mass Communication Theory, 5th edition: Sage Publications; 2000.
28. Wei D, Chan LS, Du N, Hu X, Huang YT. Gratification and its associations with problematic internet use: a systematic review and meta-analysis using use and gratification theory. Addict Behav 2024 Aug;155:108044. [doi: 10.1016/j.addbeh.2024.108044] [Medline: 38663155]
29. Bhatiasevi V. The uses and gratifications of social media and their impact on social relationships and psychological well-being. Front Psychiatry 2024;15:1260565. [doi: 10.3389/fpsyt.2024.1260565] [Medline: 38501079]
30. Froget JRL, Baghestan AG, Asfaranjan YS. A uses and gratification perspective on social media usage and online marketing. Middle East J Sci Res 2013;15:134-145. [doi: 10.5829/idosi.mejsr.2013.15.1.2127]
31. Balakrishnan P, Kroiss C, Keskes T, Friedrich B. Perception and use of reversible contraceptive methods in Germany: a social listening analysis. Womens Health (Lond) 2023;19:17455057221147390. [doi: 10.1177/17455057221147390] [Medline: 36642972]
32. Callarisa-Fiol LJ, Moliner-Tena M, Rodríguez-Artola R, Sánchez-García J. Entrepreneurship innovation using social robots in tourism: a social listening study. Rev Manag Sci 2023 Nov;17(8):2945-2971. [doi: 10.1007/s11846-023-00646-9]
33. Cole-Lewis H, Pugatch J, Sanders A, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015 Oct 27;17(10):e243. [doi: 10.2196/jmir.4969] [Medline: 26508089]
34. Pharma's digital ad spending set for a leap in 2020 as TV flattens out, forecast says. FiercePharma. URL: https://www.fiercepharma.com/marketing/annual-ad-spending-study-predicts-booming-digital-flat-tv-2020-for-pharma-and-healthcare [accessed 2026-05-29]
35. The top 10 ad spenders in big pharma for 2020. Fierce Pharma. URL: https://www.fiercepharma.com/special-report/top-10-ad-spenders-big-pharma-for-2020 [accessed 2026-05-29]
36. Top 10 pharma TV ad spenders in 2022 — anti-inflammatories and diabetes drugs dominate. Xtalks. URL: https://xtalks.com/top-10-pharma-tv-ad-spenders-in-2022-anti-inflammatories-and-diabetes-drugs-dominate-3357/ [accessed 2026-05-29]
37. Strauss A, Corbin J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques: Sage Publications; 1990.
38. Social media sentiment analysis: benefits and guide for 2026. Sprout Social. URL: https://sproutsocial.com/insights/social-media-sentiment-analysis/ [accessed 2026-05-29]
39. A marketer's guide to natural language processing (NLP). Sprout Social. URL: https://sproutsocial.com/insights/natural-language-processing/ [accessed 2026-05-29]
40. McGuirk M. Performing social media analytics with Brandwatch for Classrooms: a platform review. J Market Anal 2021;9(4):363-378. [doi: 10.1057/s41270-021-00128-5]
41. Biosimilars basics for patients. US Food and Drug Administration (FDA). URL: https://www.fda.gov/drugs/biosimilars/biosimilars-basics-patients [accessed 2026-05-29]
42. Macias W, Pashupati K, Lewis LS. A wonderful life or diarrhea and dry mouth? Policy issues of direct-to-consumer drug advertising on television. Health Commun 2007;22(3):241-252. [doi: 10.1080/10410230701626893] [Medline: 17967146]
43. Sullivan HW, Aikin KJ, Poehlman J. Communicating risk information in direct-to-consumer prescription drug television ads: a content analysis. Health Commun 2019 Feb;34(2):212-219. [doi: 10.1080/10410236.2017.1399509] [Medline: 29125325]
44. Huh J, Langteau R. Presumed influence of direct-to-consumer (DTC) prescription drug advertising on patients: the physician's perspective. J Advert 2007;36(3):151-172. [doi: 10.2753/JOA0091-3367360312]
45. Jisu H, Delorme DE, Reid LN. Perceived third-person effects and consumer attitudes on prevetting and banning DTC advertising. J Consum Aff 2006;40(1):90-116. [doi: 10.1111/j.1745-6606.2006.00047.x]

Abbreviations

DTC: direct-to-consumer
FDA: Food and Drug Administration

XSL•FO
RenderX

Edited by E Mensah; submitted 10.Oct.2025; peer-reviewed by BL Rollins, S Zwier; revised version received 04.Mar.2026; accepted 12.May.2026; published 18.Jun.2026.

Please cite as:
Macias W, Navarro Castillo MB
Tweets Surrounding Pharmaceutical Drug Brands With Top Direct-to-Consumer TV-Advertising Budgets: Social Media Listening Study
Online J Public Health Inform 2026;18:e85641
URL: https://ojphi.jmir.org/2026/1/e85641
doi:10.2196/85641

© Wendy Macias, María Belén Navarro Castillo. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 18.Jun.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Demonstrating a Social Intelligence Analysis Framework for Loneliness: Infodemiology Approach

Hurmat Ali Shah[1]; Mowafa Househ[2], PhD; Loulwah Alsumait[3], PhD; Altaf Alfarhan[3], MSc(CE)

[1]The School of Computing and Engineering, Bournemouth University, Poole, United Kingdom

[2]College of Science and Engineering, Hamad bin Khalifa University, Doha, Qatar

[3]Department of Information Science, Kuwait University, Sabah Alsalem University City, Kuwait

Corresponding Author:
Hurmat Ali Shah
The School of Computing and Engineering
Bournemouth University
Fern Barrow
Poole, BH12 5BB
United Kingdom
Phone: 44 1202524111
Email: hurmata.shah@gmail.com

Abstract

Background: Loneliness is a dynamic phenomenon that can be investigated using social media and web data.

Objective: This study aims to introduce a framework for studying loneliness through social media and online data sources. A case study is presented to demonstrate the deployment of this framework and its effectiveness in collecting and analyzing data related to loneliness.

Methods: Our proposed framework involves collecting data from various social media and online sources. We discuss the modalities of analyzing the collected data based on the framework's defined purpose. The analysis was conducted using tools such as Google Trends, the News application programming interface, X (formerly known as Twitter), Reddit, and other social media platforms. Different types of data were categorized according to the proposed framework to understand and study loneliness comprehensively.

Results: The results demonstrate the effectiveness of our proposed framework in collecting various types of data related to loneliness. Tools such as Google Trends and the News application programming interface provided insights into loneliness trends in specific regions. Social media platforms offered behavioral data on loneliness, which were analyzed using sentiment analysis and social intelligence techniques. Correlations between loneliness and personal-emotional and socioeconomic categories were identified through this analysis.

Conclusions: The framework and tools discussed in this paper complement psychosocial approaches to loneliness, which typically rely on self-report measurements. By incorporating online data perspectives, our framework provides valuable insights into loneliness dynamics, enhancing our understanding of this complex phenomenon.

(Online J Public Health Inform 2026;18:e59861) doi:10.2196/59861

KEYWORDS

health informatics; loneliness informatics; loneliness theory; health effects; loneliness interventions; ICT-based interventions; social media–based interventions; social media; ICT; lonely; loneliness; social isolation; analysis framework; Twitter; Reddit; behavioral data

Introduction

Background

Loneliness has global public health consequences. Loneliness not only affects the mental health of people worldwide but also has consequences for physical health [1]. Loneliness is a dynamic phenomenon that is understood from multiple perspectives and disciplines [2]. It can be studied from an information and health informatics perspective. Applying data and data science disciplines to study health in rapidly changing scenarios has led to the development of fields such as infodemiology and infoveillance [3]. There is a demand for a

XSL•FO
RenderX

framework that is based on the tools of infodemiology to study loneliness through data sources available online and from social media.

The proposed social intelligence analysis framework for loneliness in this paper has four parts: (1) identifying trends, (2) monitoring the news, (3) exploring the breadth of topics, and (4) finally analyzing the depth of topics. In the first stage, it is important to know how the phenomenon being investigated has a general trend. This stage gives us the overall scope of the topic and its temporal dimensions. When the temporal dimensions are known, we can go to the second stage of analysis, which is to know whether the phenomenon is getting coverage in a specific geographical area. The third stage of the analysis is to use social media data to analyze the correlations and associations of the phenomenon in the geographical area. This part can focus on the breadth or the variety of related topics and correlations. The fourth stage, which can compound the third stage, is to provide details on the topics and correlations of the phenomenon.

This paper aims to provide a demonstration through data collection and data analysis according to the proposed framework. Social media and online sources can help us understand the prevalence of loneliness to devise technology-based and community-oriented strategies to address it. While technology may have resulted in a fragmented and individualized existence, it can also be a great healer. The rise of social media has transformed the way in which we interact with others, offering new opportunities for social connection and communication. Loneliness is a common experience that can have negative effects on mental and physical health, and social media use has been implicated as a potential contributor to loneliness [4]. Governments such as Japan and the United Kingdom have designated positions dedicated to loneliness. In response to rising concerns about social isolation, particularly among older adults and young people, Japan appointed a Minister of Loneliness in 2021 [5]. As can be seen, in addition to piquing the interest of scholars, with the engagement of governments, loneliness has become a component of public health.

Objectives

The proposed framework uses Google Trends, the News application programming interface (API), and data from X (formerly known as Twitter) and Reddit under the interdisciplinary field of infodemiology. Further studies and discussions in infodemiology can be found in the works by Jia et al [6], Eysenbach [7], and Yu et al [8]. We make a distinction between web and social media sources because social media sources are self-reported and can provide an intimate and personal perspective. By web sources, we mean sources other than social media. Although the main focus of this work was on the United States, country-specific filtering can be used for Google Trends, the News APIs, and X data. The Reddit API does not provide the data for a particular country, so Reddit data on loneliness only includes worldwide posts. This is one of the limitations of this demonstration of the social intelligence analysis framework. Nonetheless, Reddit data can still provide useful insights for the study of loneliness. We used the sentiment intensity analyzer contained in the Natural Language Toolkit (NLTK; Team NLTK) and Valence Aware Dictionary and Sentiment Reasoner (VADER) [9] from the NLTK for various analyses in this study.

This study aims to introduce a framework for studying loneliness through social media and online data sources. The framework is important to understand loneliness using data available online and to complement the theoretical and psychosocial understanding of loneliness.

Methods

Overview

Most researchers in the fields of sociology, public health, and psychology have studied loneliness using the University of California, Los Angeles (UCLA), Loneliness Scale [10-12]. The UCLA Loneliness Scale is a measuring instrument developed by Russell [13] at UCLA. It is an essential instrument for assessing subjective perceptions of loneliness. The scale comprises 20 items. The UCLA Loneliness Scale investigates various dimensions of loneliness involving social isolation, relational quality, and self-reliance. Its core domains—social connectedness, relational connectedness, and self-reliance—investigate the availability and depth of social interactions and assess an individual's capacity to manage loneliness. It has been broadly used in psychological research, specifically in assessing the effects of loneliness on mental health and social behaviors across diverse demographic groups. The UCLA Loneliness Scale is a valued quantitative measure [13]. The proposed framework provides a complete assessment of loneliness, helping to identify, recognize, and theoretically address feelings of isolation, thereby generating discussions about social associations and guiding possible interventions to allay loneliness.

The UCLA Loneliness Scale is a set of questions, whereas our framework collects behavioral information and unstructured text data, in addition to other online data, to understand loneliness. For the sake of brevity, a detailed explanation of the proposed framework is not included in this paper.

Proposed Framework

The proposed social intelligence analysis framework for studying loneliness leverages a wide range of data sources from across the web and social media, addressing the challenges of extracting meaningful information from the overwhelming volume of available online content. Traditional measures of loneliness, such as the UCLA Loneliness Scale, have long been used in scientific and psychosocial research to assess individuals' subjective feelings of social isolation, well-being, and connection to others. However, these measures rely heavily on self-reported survey data, which while valuable, only capture loneliness in controlled, specific contexts. In contrast, the proposed framework uses real-time, publicly available online data to offer a more dynamic and expansive perspective on loneliness as it naturally occurs in society. The framework is divided into four key parts: identifying trends, following the news, analyzing the range of topics, and examining the depth of discussion.

XSL•FO
RenderX

In the first stage, identifying trends, Google Trends is used to track the frequency with which people search for loneliness-related terms over time. This tool allows for the analysis of temporal patterns in public interest, offering insights into the external factors—such as societal events, economic downturns, or health crises—that may cause fluctuations in loneliness. For example, spikes in searches for loneliness-related terms might coincide with lockdowns during the COVID-19 pandemic, indicating increased public concern. In addition, Google Trends provides regional data on where these searches originate, helping researchers and policymakers target resources and interventions to the areas most affected by loneliness. Google Trends also offers related search queries, enabling the discovery of connected terms such as "loneliness in older adults" or "loneliness and mental health," which can guide further research and exploration.

The second stage, following the news, involves analyzing news articles using news APIs, such as the News API, Google News API, and Bing News Search API. News coverage of loneliness reflects broader societal interest and how loneliness is framed and discussed in the media. By examining trends and patterns in news reporting, researchers can gain insights into the causes, consequences, and public perceptions of loneliness. Media coverage often highlights demographic variations, such as the loneliness of older adults or teenagers, and reveals how loneliness is discussed within the context of mental health, social isolation, or public health crises. News stories often feature personal experiences, providing a deeper look into how loneliness affects individuals. In addition, news analysis allows researchers to monitor how public awareness of loneliness evolves and how media framing might influence public attitudes or contribute to the stigma surrounding loneliness.

In the third stage, analyzing the range of topics, the focus shifts to social media platforms, particularly X, where users express their personal feelings and opinions in real time. Through keyword searches and sentiment analysis of X data, researchers can observe the range of experiences and emotional responses associated with loneliness. The short-form, real-time nature of posts on X allows for the collection of self-reported loneliness experiences, capturing personal, emotional, and psychological aspects of the phenomenon. Furthermore, the wide range of topics and hashtags related to loneliness can help researchers understand the broader social, economic, and political factors influencing loneliness, providing a more diverse understanding of the issue.

Finally, in the fourth stage, examining the depth of discussion, platforms such as Reddit provide a more in-depth exploration of loneliness through longer, more detailed posts and discussions. Reddit users often engage in communities, or subreddits, dedicated to specific topics, such as *r/loneliness* or *r/depression*, where they share personal experiences and seek advice. This detailed, often anonymous sharing allows for more honest and comprehensive insights into the complexities of loneliness. The depth of these discussions makes Reddit a valuable tool for uncovering the more nuanced, personal dimensions of loneliness, particularly its emotional and psychological impacts. Reddit's forum-based structure also allows researchers to track the evolution of discussions over time and identify recurring themes and subtopics, contributing to a deeper understanding of loneliness.

Demonstrating the Proposed Framework

In the initial implementation of the framework, the aim is to gain an understanding of the underlying patterns associated with the phenomenon under investigation. This initial stage provides a comprehensive view of the topic and its temporal aspects. Once these temporal dimensions are determined, we can proceed to the second stage of analysis, which involves assessing whether the phenomenon is receiving attention within specific geographic regions. The third stage of the analysis entails using social media data to explore the relationships pertaining to the phenomenon within these geographical areas. This stage can either focus on the diversity and the broad spectrum of the associated topics and correlations or delve into specific aspects.

Building on the insights gained in the third stage, the fourth stage involves a more in-depth examination of the topics and correlations associated with the phenomenon. In the following sections, we will explain each of these stages in detail, outlining the tools and methodologies that will be used to facilitate their execution.

This paper aims to demonstrate the social intelligence analysis framework through a case study in which data on loneliness were collected from online data sources. The major contributions of this paper are as follows: (1) demonstrating how data can be collected in an organized way and how to analyze them to gain meaningful insights about the nature of loneliness, (2) demonstrating how different online and social media data sources can provide varied information on the dynamic and changing nature of loneliness, and (3) categorizing the themes and topics associated with loneliness into socioeconomic and personal-emotional or other relevant categories from the data collected and processed through the social intelligence analysis framework.

Data Collection

As this framework involves four different data sources, the data collection for each data source followed the specifics of the associated API and the rules of the data source. The data sources were Google Trends, the News API, X, and Reddit. First, data from Google Trends were collected for the year 2022. The dashboard of Google Trends allows for searching for a particular country using keywords, as well as searching for a specific year. For the news analysis, we used the News API in Python (Python Software Foundation). The data were collected for the keyword "loneliness" in the United States. On the basis of the data collected in this stage, the analysis could focus exhaustively on specific cities or countries to collect more data about them. However, we did not want to limit the search to one specific country to allow for the collected data to be a proof of concept. The collected data on X were merged based on location, user ID, and post ID to identify posts from the United States. The total number of posts was 100,000. The words "lonely," "loneliness," "alone," "isolated," and "isolation" were used to retrieve the posts.

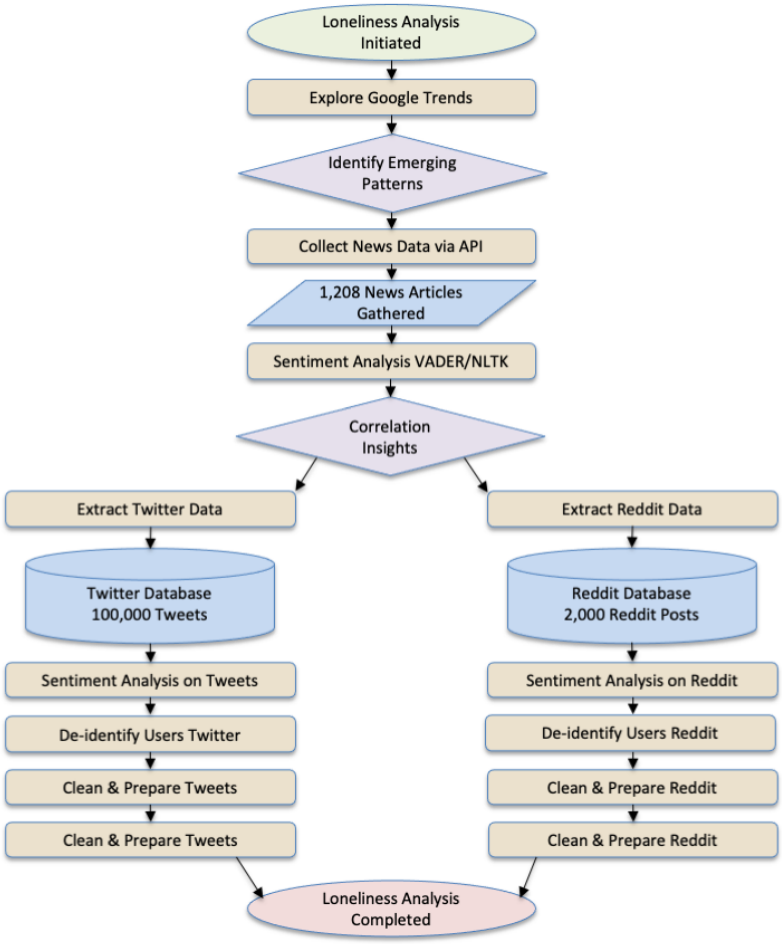
The Reddit data collection methodology is relatively straightforward. Reddit is a forum-based social media platform

XSL•FO
RenderX

where people post about a topic on a subforum dedicated to it. These subforums are called subreddits. The Reddit API provides access to individual subreddits to download the top posts on a topic, which is determined by the number of upvotes and other parameters of engagement. The posts from the *r/loneliness* subreddit were collected through the Reddit API. The *r/loneliness* subreddit has 13,000 members who can post and comment in this subforum. Reddit has its own algorithm for giving scores (ie, higher visibility to posts), which also contains input from other users in the form of upvoting.

We collected the top 2000 posts from the *r/loneliness* subreddit with all their comments. The comments varied for each post, both in number and size. It is worth noting that some of the comments were of the same length or even longer than the original posts. Thus, the comments constituted valuable data on loneliness. In total, more than 2000 individual texts were analyzed, which was estimated by multiplying the posts by the average number of comments per post. While some posts did not have comments, the maximum number of comments for a single post was 55. The average number of comments was 4.51, and the total number of comments was 8570. When combined with the posts, this resulted in more than 10,000 unique texts or personal expressions of loneliness from Reddit. We analyzed both the posts and the comments to determine the frequency of occurrence of words to locate the correlations of topics and themes with loneliness. The flowchart in Figure 1 shows the details of the data collection and analysis process used in this study.

Figure 1. Loneliness framework flowchart. API: application programming interface; NLTK: Natural Language Toolkit; VADER: Valence Aware Dictionary and Sentiment Reasoner.



Data Analysis

X Data Analysis

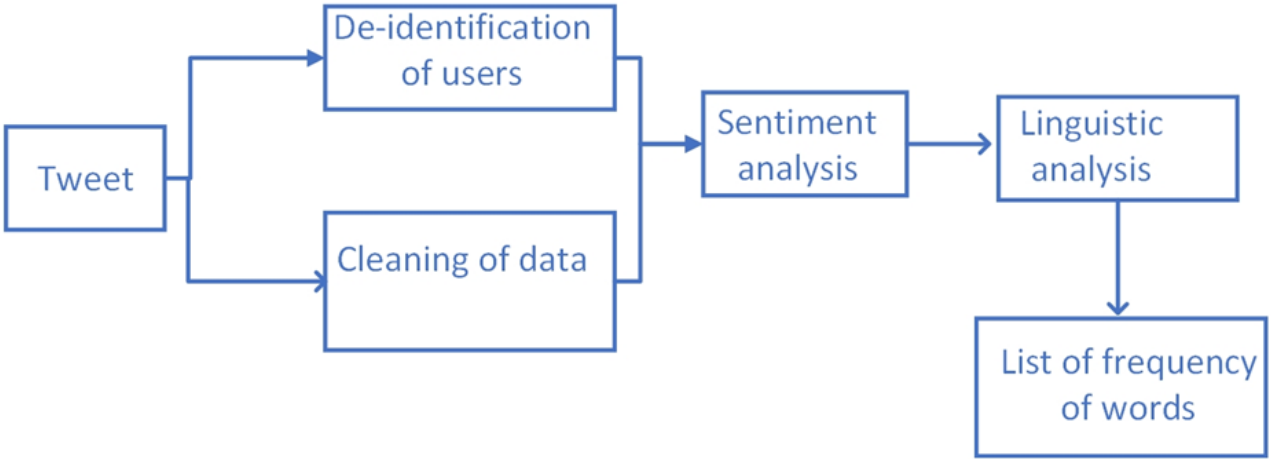
We collected a particular number of X posts using keywords related to loneliness. If we reported all the X posts that contained feelings of loneliness, we would not have required a further stage, but the question here is how the expression of loneliness can imply negative consequences, such as mental health problems. In that case, the problem becomes determining the association or correlation between themes (which may represent loneliness) and keywords representing loneliness. For instance, we had to determine the relationship between “hurt,” “sick,” “tired,” and “sleep” and the expression of loneliness. This task is usually carried out by associating lexicon categories with posts including the words “lonely” or “alone.”

The problem we formulate in this paper is broader in scale. Thus, the limited scale of representative X posts had to be interpreted in a novel way to provide meaningful insight into loneliness. All the posts in the dataset contained keywords representing loneliness. These data could be analyzed to find the association between loneliness and other socioeconomic or personal-emotional categories worldwide or for individual

XSL•FO
RenderX

countries. Analyzing these data is important to provide a global picture of the determinants of loneliness and to provide a tool for policymakers to address loneliness in their specific countries. However, "lonely" or "alone" can also be mentioned in a nonnegative way. Using sentiment analysis and manual analysis of the topic and themes of negative posts allowed us to look at the relationship between mentioning keywords representing loneliness and negative emotions, which may ultimately be linked to psycholinguistic features of mental well-being (Figure 2).

Figure 2. Pipeline for processing Twitter data.



Loneliness in the context of mental health is a negative emotion, which is why the sentiment analysis stage is required—to find out how loneliness is expressed. For an analysis of loneliness in the context of mental health, we filtered out the X posts in which the expression of loneliness was negative. The collected posts also contained metaphorical uses of "lonely" or "loneliness" that did not pertain to our use of loneliness. Such mentions of loneliness were present in positive- and neutral-sentiment posts. The definition of loneliness in this paper connotes a negative feeling. While loneliness can also be a positive or neutral feeling for some people or at certain times, when it comes to its association with mental health issues, the negative consequences of loneliness must be considered.

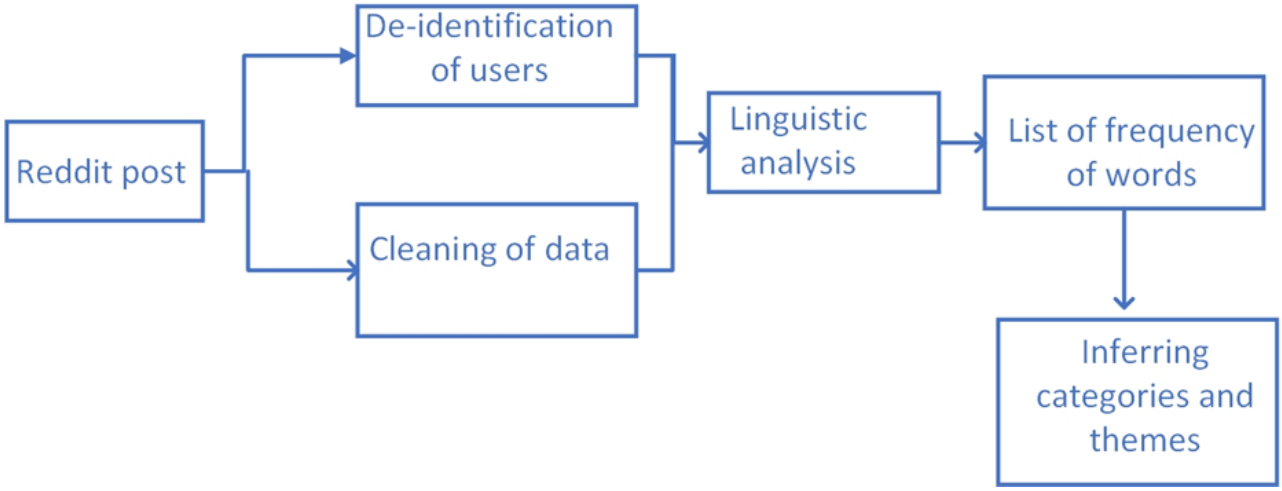
We conducted sentiment analysis on both news articles and X data. The news articles were analyzed using the sentiment intensity analyzer contained in the NLTK. The collected posts were stored in a database, and a sentiment analysis was conducted using VADER from the NLTK. VADER is a lexicon and rule-based model for sentiment analysis. The lexicon-based algorithm is constructed using a dictionary that contains a detailed list of sentiment features. In addition, VADER complements the lexicon-based dictionary with grammatical rules that are heuristic in nature and used to determine the polarity of the sentiment. The resulting polarity of the sentiment analysis was used as an indication of loneliness in the dataset. For the sake of brevity, we will not go into the details of using VADER and sentiment analysis. For interested readers, we recommend referring to our previous work [14-16].

Reddit Data Analysis

Figure 3 shows the pipeline for processing Reddit posts. The difference between Figures 2 and 3 is the absence of sentiment analysis on the Reddit posts. After going over the subreddit *r/loneliness*, we found that the posts were about the emotional expression of loneliness and did not involve metaphorical or non-sequitur uses of "loneliness." Reddit and its subreddits are characterized by serious engagement on the topics that the subreddits are designed for. Therefore, no sentiment analysis of the Reddit data was deemed important, and the posts were analyzed using the frequency of occurrence of words to find out the themes and topics that were most highly associated with loneliness.

XSL•FO
RenderX

Figure 3. Pipeline to process Reddit data.



Manual Coding and Analysis of X and Reddit Data

The authors cleaned the data before analysis. We ensured that the X posts were deidentified by removing usernames and IDs as part of the data cleaning process. While the data are publicly available, we did not disclose any collected data without first anonymizing them. Sentiment analysis was conducted after cleaning the data, which included removing redundant characters, numbers, special characters, users' profile IDs, and information such as reposts. For the Reddit data, direct analysis was possible. However, posts from bots or other automated and potentially malicious agents were not filtered out in this study, a limitation that we plan to address in future work by removing such posts before analysis.

We stored X posts with a negative sentiment separately for further analysis, focusing on identifying prominent themes and categories through manual coding. After removing stop words and applying lemmatization to reduce word count, we generated a compact list of word occurrences. This list was manually analyzed to identify larger socioeconomic or emotional-personal categories guided by the literature, although the process remained subjective, relying on the researchers' judgment. For Reddit data, we followed a similar process, collecting posts and comments, removing stop words, applying lemmatization, and generating a word occurrence list for analysis without conducting sentiment analysis on the data from this platform.

Manual coding and analysis were used to assess expressions of loneliness on X and Reddit objectively. This topic-based categorization was more effective in identifying meaningful similarities and differences. Unlike the n-gram method, which focuses on word co-occurrence, our inductive approach allowed themes to emerge organically, providing a thorough analysis without being constrained by predefined keywords. This method, being quantitative, avoids subjective interpretation, relying instead on the frequency of word occurrences and their classification into relevant categories grounded in existing literature. The detailed analysis method and the use of sentiment analysis for Reddit and X data can be found in our previous work [14-16].

News and Google Trends Analysis

The methodology used in this study involved using the News API tool for data analysis. The News API provides programmatic access to a vast collection of news articles from various sources. The data analysis process began by formulating relevant search queries and parameters to retrieve news articles specifically related to loneliness. These parameters included keywords such as "loneliness." The News API facilitates the retrieval of a significant volume of news articles encompassing different geographical regions and periods. The collected data underwent preprocessing, including cleaning, filtering, and removing duplicate or irrelevant articles. Subsequently, sentiment analysis was used on the news articles. Sentiment analysis for news articles was used for the same reasoning explained previously for the analysis of X posts. These analyses aimed to identify prevalent themes, trends, and sentiments associated with loneliness.

Google Trends provides access to a vast database of search queries and allows for the analysis of search interest over time and across different regions. The data analysis process for Google Trends began by selecting relevant keywords related to loneliness. These keywords were used to retrieve search interest data from Google Trends. The retrieved data were then processed and analyzed to identify temporal patterns, regional variations, and related queries associated with loneliness. The analysis involved examining trend graphs, comparing search interests across different regions, and identifying related topics and queries.

Ethical Considerations

All data such as usernames, tweets, quotes, etc, in the paper have been deidentified.

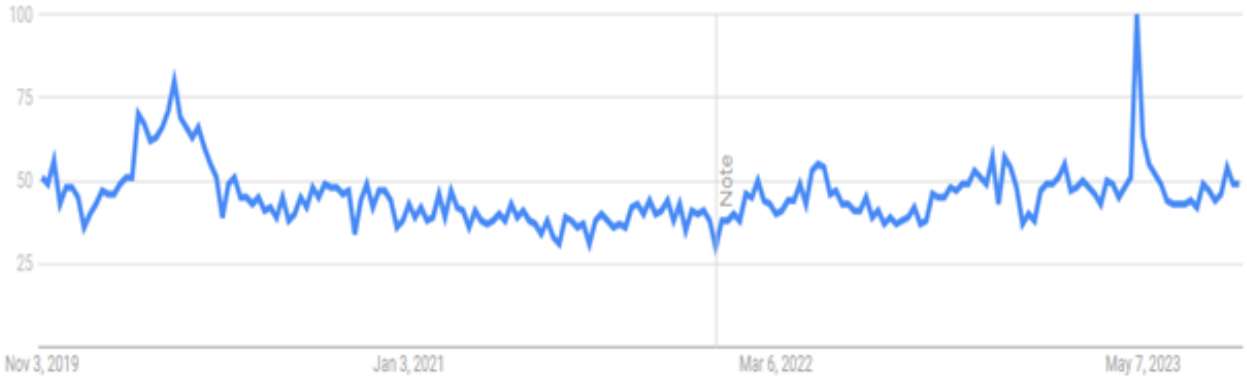
Results

As the first stage involved knowing the trends, we carried out a search for the term "loneliness" on Google Trends, shown in Figure 4. We selected a longer period starting before the COVID-19 pandemic, specifically from November 1, 2019, to August 31, 2023. Figure 4 shows a snapshot of the trend graph for "loneliness." The "Note" breakpoint in the graph represents

XSL•FO
RenderX

the improvement to Google's data collection system on January 1, 2022. The y-axis represents interest over time in the topic. A value of 100 represents peak interest in and popularity of the topic, whereas a value of 50 means that the term had half the popularity. The data points were collected weekly. There was a peak in interest in the topic on May 7, 2023, which did not correspond to a particular event and seems to be an outlier or an anomaly. On the other hand, the interest in the topic was at higher levels during the months of lockdowns related to the COVID-19 pandemic, peaking around the end of March 2020. Overall, the graph shows that the interest in loneliness remained at approximately half the peak levels throughout this period. This shows a sustained interest in the topic. Although the number of searches for the term and its volume are not provided, the popularity rates provide an insight into the trends for loneliness (or any other term) over time.

Figure 4. Google Trends chart for the term "loneliness." The "Note" breakpoint in the graph represents the improvement to Google's data collection system on January 1, 2022.



Google Trends also provides *Related queries* for the topic. In case of the search term "loneliness," the related queries were "covid loneliness," "loneliness during covid," "my loneliness is killing me tiktok," "is the cure to male loneliness," and "surgeon general loneliness epidemic." These terms can express different socioeconomic, personal-emotional, or other phenomena associated with loneliness. If further insight into loneliness is required, these terms can be searched separately, and the results can be compared. Google Trends also provides a tool in which two different topics or queries can be searched.

In the second stage, following the news, we used the News API in Python to retrieve news articles containing mentions of loneliness. In total, we retrieved 956 articles. Table 1 includes a random selection of 25 articles. We carried out a sentiment analysis of the news articles retrieved. An overall negative sentiment score means that the article discussed topics or themes that were negatively associated with either loneliness or broader mental health issues. The news articles with negative sentiment scores can be read for further trend analysis.

XSL•FO
RenderX

Table 1. A list of news articles with their sentiment analysis scores.

Article title	Sentiment score
Meet The People Who Listen to Podcasts Crazy-Fast	−0.766
MORABITO: Hillary Clinton Just Gave Away the Left's playbook for censorship and oppression	0.681
The Connection Cure: 6 Ways to Beat Loneliness	−0.661
Official Trailer for Babak Jalali's 'Fremont'	0.077
4 Signs Trauma Has Affected Your Self-Worth	−0.944
Why Historian Jill Lepore Hated Barbie	−0.166
MJ Lenderman Nods to Bob Dylan on New Single "Knockin"	0.700
How Athletic Beer Won Over America	0.215
4 ways simulation training alleviates team burnout	−0.681
Nessa Barrett ON: How to Overcome Loneliness	−0.851
5 Ways Men Can Build Strong Connections	0.971
Gywneth Paltrow saw you from across the bar and wants you to stay with her	−0.296
Album Of The Week: Ratboys The Window	0.250
What is the 'Joy' in the Joy of Missing	0.212
Self checkout could be making Americans Lonelier	0.772
Leave it to the dogs (13 Photos)	−0.700
An Easy Way to Reduce Depression And Loneliness	−0.968
Perils of not being attractive or athletic	−0.908
Parents Are Almost as Depressed and Anxious as Teens	−0.900
Bike Happy Hour, listening, and loneliness	0.908
3 Ways Teachers Can Instill Belonging in Students	0.898
Let It Be Sunday, 325!	0.338
How to Overcome Feeling Lonely and Powerless	−0.953
Edinburgh Fringe: The Life and Times of Michael K	0.869

Another analysis that can be carried out on the collected news articles is a list of bigrams in collocations. A collocation is a series of words that co-occur more often than would be determined by chance. In Textbox 1, we collected the bigram collocations (ie, a combination of 2 words that occurred together in the collected news articles). Although the list is a small sample and contains words that may connote difficulties regarding loneliness, collecting bigrams in collocations can provide a wider impression of what themes and topics are discussed in conjunction with loneliness. This, in turn, can point to other directions for exploring the dynamics of loneliness.

XSL•FO
RenderX

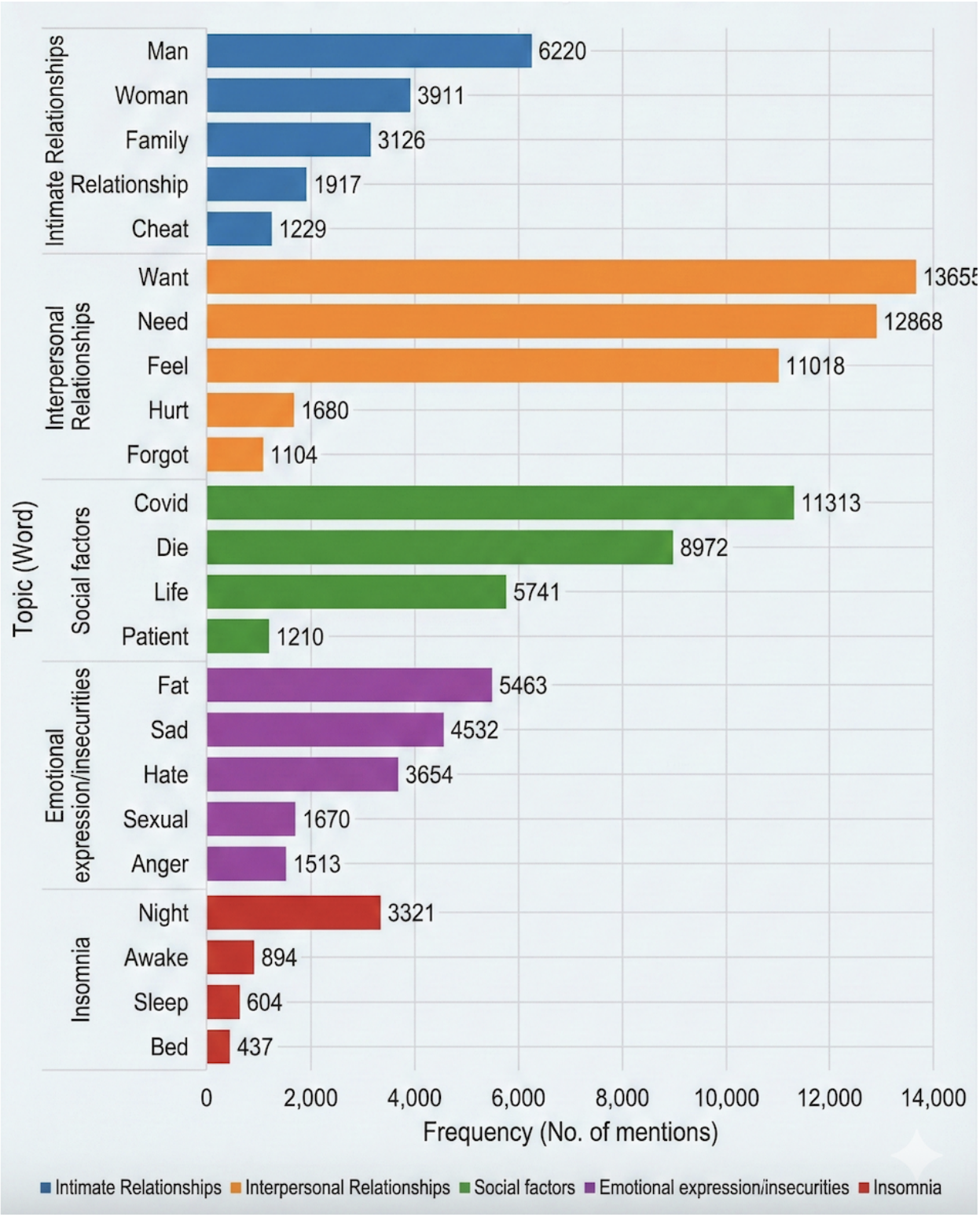
Textbox 1. Bigram collocations.

- “Relationship” and “loneliness”
- “Insomnia” and “symptoms”
- “Loneliness” and “purpose”
- “Depression” and “bitterness”
- “Anxiety” and “disorder”
- “Isolation” and “silence”
- “Alcohol” and “misuse”
- “Anxiety” and “loneliness”
- “Epidemics” and “obesity”
- “Loneliness” and “long”
- “Help” and “loneliness”
- “Insomnia” and “symptoms”

For stage 3, the analysis of the range of topics and topic analysis was conducted on the X posts. Table 2 shows the results of relevant themes and categories from analyzing the word occurrence in posts, whereas Figure 5 shows a visualization of the most dominant themes. We carried out sentiment analysis on 200,000 posts and found that 30.7% (n=61,400) had a negative sentiment. Table 2 breaks down the text of these negative-sentiment posts into the resultant words. Posts containing the keywords mentioned in the Methods section were collected. Sentiment analysis was then carried out. Sentiment analysis differentiates between phrases and topics that carry meaningful information on loneliness and those that use the term in a metaphorical or non-sequitur manner.

XSL•FO
RenderX

Figure 5. Visualization of highly correlated themes obtained from X posts.



XSL•FO
RenderX

Table 2. Words highly correlated with mentions of loneliness in X posts. Topics are categorized under a broader thematic area.

Thematic area and topic	Mentions, n
Intimate relationships	
"Cheat"	1229
"Man"	6220
"Family"	3126
"Woman"	3911
"Relationship"	1917
Interpersonal relationships	
"Want"	13,655
"Need"	12,868
"Feel"	11,018
"Hurt"	1680
"Forgot"	1104
Social factors	
"Covid"	11,313
"Die"	8972
"Life"	5741
"Patient"	1210
Emotional expressions or insecurities	
"Sad"	4532
"Hate"	3654
"Fat"	5463
"Anger"	1513
"Sexual"	1670
Insomnia	
"Night"	3321
"Awake"	894
"Sleep"	604
"Bed"	437

The results show that most of the X posts containing keywords associated with loneliness from the United States were neutral, which means that they did not meaningfully contribute to the analysis of loneliness. Before conducting the detailed analysis of the posts on loneliness, it was important to identify uses of "loneliness" as a metaphor or non sequitur (ie, those posts that would not add meaningfully to the analysis of negative consequences related to loneliness). Neutrality can also represent the mention of loneliness in descriptive terms.

The basic analysis of the Reddit data for stage 4, examining the depth of the discussions, is provided in Table 3. We collected the top 2000 Reddit posts from the *r/loneliness* subreddit with all their comments. Thus, we analyzed more than 2000 total individual texts. The breakup of the data into words resulted in more than 25,000 words. For the sake of meaningful mentions of topics and brevity, we set a threshold of 50 topics that gave us 411 words to be analyzed. It should be noted that a significant number of these words were language constructs. Only the words that were meaningful in terms of emotions or other expressive qualities were included in the analysis.

Table 3. Analysis of frequency of occurrence of words in the Reddit data (N=35,057).

	Words, n (%)
Words occurring >100 times	611 (1.74)
Words occurring >1000 times	78 (0.22)

XSL•FO
RenderX

In addition, we want to note that the Reddit posts on loneliness were not specific to the United States. The posts were not divided by country, and the Reddit API does not allow for country-specific downloads. Some methods provide the ability to find the country of the post from the Reddit data, but this involves processes that are beyond the scope of this paper [17]. Table 4 and Figure 6 list and visualize the correlations between themes and loneliness in the *r/loneliness* subreddit. It can be observed from the table that the themes are mostly focused on relations and emotional expression. Because of the longer posts, it is expected that people would have more space to open up and express their feelings. Social media platforms such as Reddit provide spaces where individuals can express their vulnerabilities without facing backlash that can come in the form of social ostracization.

XSL•FO
RenderX

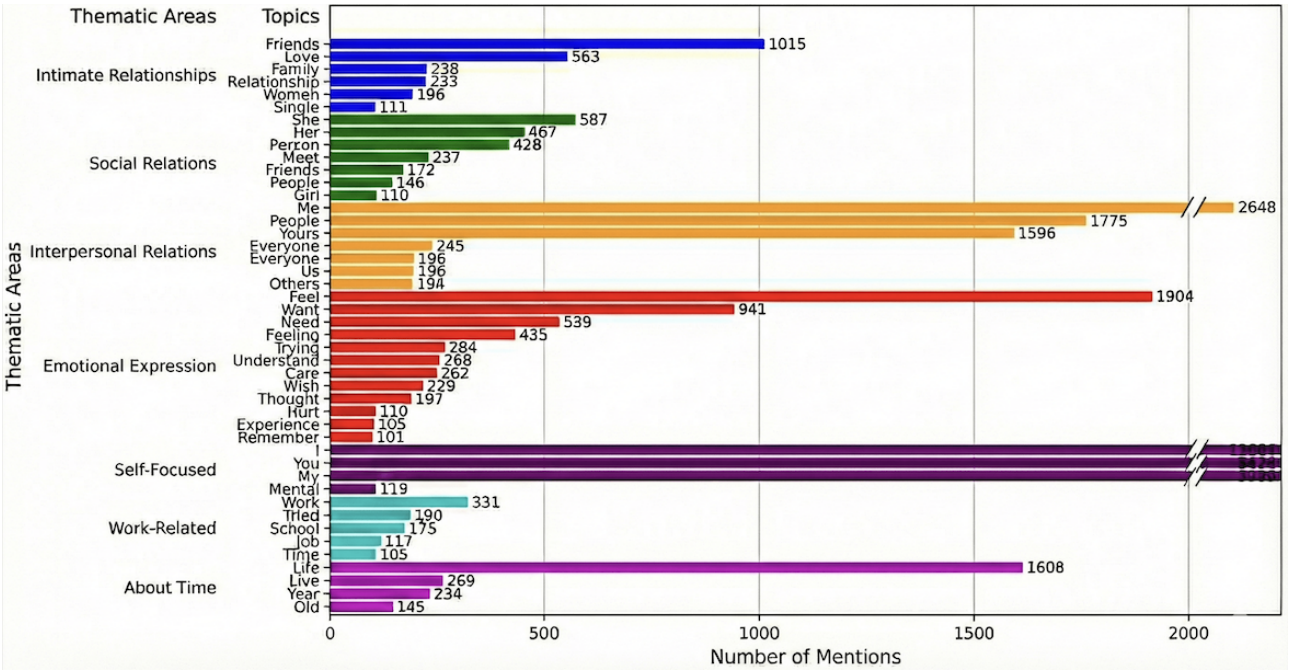
Table 4. Correlations of themes with loneliness in the Reddit data. Topics are categorized under a broader thematic area.

Thematic area and topic	Mentions, n
Intimate relationships	
“Love”	563
“Women”	196
“Relationship”	233
“Family”	238
“Single”	111
“Friends”	1015
Social relations	
“Friends”	172
“Girl”	110
“She”	587
“Her”	467
“People”	146
“Online”	106
“Meet”	237
“Person”	428
Interpersonal relations	
“Me”	2648
“He”	101
“Yours”	1596
“Us”	196
“Everyone”	245
“People”	1775
“Others”	194
Emotional expression	
“Thought”	197
“Hurt”	110
“Trying”	284
“Pain”	114
“Experience”	105
“Remember”	101
“Understand”	268
“Feeling”	435
“Want”	941
“Need”	539
“Feel”	1904
“Wish”	229
“Care”	262
Self-focused	
“I”	13,604
“Mental”	119
“My”	3989

XSL•FO
RenderX

Thematic area and topic	Mentions, n
"You"	5424
Work related	
"Work"	331
"Job"	117
"Tried"	190
"Time"	105
"School"	175
Time related	
"Life"	1608
"Year"	234
"Live"	269
"Old"	145

Figure 6. Visualization of highly correlated themes obtained from Reddit posts.



Discussion

Principal Findings

This paper demonstrates our social intelligence analysis framework for studying loneliness from online and social media data sources, and presents an overall picture of how a varied topic such as loneliness can benefit from multiple levels of analysis. In this study, we adopted a comprehensive approach, integrating data from Google Trends, news articles, X, and Reddit to examine the multifaceted concept of loneliness within the framework of social intelligence analysis. The demonstration of the social intelligence analysis framework for loneliness revealed interesting patterns, such as in Google Trends, and provided the topics related to mentions of loneliness.

Analysis of Google Trends data exposed intriguing temporal patterns in the public's interest in loneliness. We observed notable spikes in loneliness-related search queries at various junctures, suggesting that external events, cultural shifts, or seasonal influences may significantly impact the prevalence and perception of loneliness in society. Our examination of news articles provided a broader contextual understanding of loneliness. The sentiment analysis of news articles provided a helpful tool to gather news articles that discuss the negative and health consequences of loneliness.

From a psychological perspective, the increases in the Google Trends graph indicate elevated public interest in loneliness that can be explained as societal reactions to noteworthy or unexpected events [13]. These events, such as the COVID-19 pandemic and the social and physical restrictions that followed, made people feel more emotionally and psychologically alone. Our sentiment analysis of news items revealed that media coverage frequently reflects this elevated awareness. In addition to reporting on these occurrences, the media also influences

XSL•FO
RenderX

public opinion by highlighting the psychological results of loneliness, particularly its detrimental consequences on mental health. This combination of social factors and psychological reactions, as observed in media coverage and Google Trends, highlights the multifaceted nature of loneliness [3].

The analysis of X posts and Reddit posts revealed associations between socioeconomic and personal-emotional factors and loneliness. These factors included emotion, sentiment, emojis, and topic modeling. This analysis demonstrated that such factors could help gather evidence and analyze interactions on the topic of loneliness and other related topics. The first factor was emotion, which can serve as a guide in understanding people's reactions. The second most common factor was relationships. Other thematic areas such as health, work, self-focused topics, and insomnia-related topics indicate the intimate nature of loneliness.

The difference that was observed between the data from X and Reddit (ie, stage 3 and stage 4 of the framework) was in their diversity and extensiveness. In the X data, a range or diversity of topics and themes could be observed. Because of the limited character expression on X, people express their thoughts or opinions in a compact manner; however, through analysis of the terms used and the overall sentiment of the sentences, an association with loneliness can be found. There can be a range of such themes in which there are direct mentions of loneliness in a negative context. On the other hand, Reddit data can be useful for finding the depth of a theme associated with loneliness (ie, what subthemes or topics under a broader category are related to loneliness). These data are important for investigating the possible causes of loneliness. The diversity of the discovered topics and themes from X and the depth of topics that were found on Reddit can be used in complementary ways.

The framework delineated in this paper provides a versatile, multistep approach to analyzing loneliness through online and social media data. Beyond studying loneliness, this framework can be expanded to explore other complex societal issues, such as mental health conditions (eg, anxiety and depression), misinformation, or public reactions to crises. In addition, it can be used for early detection of public health trends or social phenomena by monitoring real-time data. The framework's capacity for sentiment analysis and topic modeling can offer valuable insights into emotional and psychological responses, which can be applied to develop targeted interventions, inform policies, or enhance public health programs.

The results of this framework reveal the complex, multifaceted nature of loneliness, highlighting its emotional, psychological, and socioeconomic dimensions. These insights can be used in mental health applications by enabling early identification of loneliness trends and allowing for real-time monitoring of at-risk groups. For mental health patient care, these data can be integrated into artificial intelligence–driven tools that personalize interventions, offer resources, or connect patients with support networks. It can also help inform health care providers about socioenvironmental triggers contributing to loneliness. Future research can incorporate more advanced natural language processing tools and extend the use of this framework to cross-cultural studies, improving understanding of how societal factors impact loneliness and other issues across different populations.

The proposed framework can be used in future research endeavors to deepen the understanding of loneliness and its societal implications by providing a systematic approach to analyze diverse and large-scale data from online platforms. By capturing both temporal trends and geographic relevance, researchers can identify key moments and regions where loneliness spikes, enabling a more focused examination of societal or environmental triggers. Expanding the framework to include more detailed demographic information will allow researchers to study how loneliness impacts specific groups, such as older adults or the younger generations, across various cultural contexts. In addition, the framework's ability to integrate multiple data sources, including social media platforms, news articles, and search trends, offers a more holistic perspective of how loneliness is discussed and experienced at both personal and collective levels. This could lead to a deeper exploration of the role that socioeconomic factors, public health crises, or policy changes play in exacerbating or alleviating loneliness. Furthermore, the sentiment and thematic analysis components can be refined to investigate emotional undercurrents related to loneliness, helping uncover the psychological and emotional dimensions of social isolation.

This framework can support the development of artificial intelligence–driven tools for real-time monitoring and intervention, ultimately informing policy and community-based solutions to address loneliness more effectively. The proposed framework could be adapted to investigate various other societal and public health issues that are influenced by dynamic social and environmental factors. For instance, mental health conditions such as anxiety and depression, which often correlate with loneliness, could be explored by tracking online discourse, sentiment, and search patterns. The framework could also be applied to study the societal impacts of major events such as pandemics, economic downturns, or political crises, where real-time social media analysis could provide insights into public emotions, coping mechanisms, and socioeconomic concerns. In addition, issues such as misinformation, public perceptions of health interventions, or even social phenomena such as digital addiction or climate anxiety could be investigated. By analyzing data from different online platforms, researchers can gain a more comprehensive understanding of public reactions and trends related to these complex, evolving issues.

Limitations

While our research yielded valuable insights into loneliness using an innovative approach, there are also some limitations. First, our reliance on digital data sources such as X and Reddit may introduce biases. These platforms primarily represent individuals comfortable with sharing their experiences online, potentially excluding those who are less active or lack internet access. This is a limitation of the proposed framework that can be overcome through in-depth interviews or surveys to provide a more holistic understanding of individuals' emotions, motivations, and coping mechanisms.

In addition, the study's temporal analysis of Google Trends data lacks causality. While we identified spikes in search queries,

XSL•FO
RenderX

determining the specific reasons behind these fluctuations requires further investigation. Furthermore, the focus of the study on English-language data may not fully capture the global diversity of loneliness experiences, potentially limiting the generalizability of our findings. Another limitation lies in the demonstration itself, which relied on data from Reddit, in which the country cannot be specified. For a nuanced understanding of data from Reddit, the data first need to be categorized by region.

Conclusions

In this paper, we introduced a comprehensive framework for analyzing loneliness through the lens of social intelligence analysis. The framework uses data from diverse online sources, including search engines, news articles, X, and forum websites. This paper provides a demonstration of our proposed framework and reveals correlations between loneliness and online news and posts through sentiment analysis. We provided details on how data can be collected and analyzed according to the umbrella of our proposed framework for studying loneliness through social media and online data. In addition, sentiment analysis of news articles sheds light on the negative health consequences of loneliness, whereas the analysis of X posts and Reddit posts revealed associations between loneliness and various socioeconomic and personal-emotional factors.

Despite the framework limitations, our study provides valuable insights into the multifaceted nature of loneliness through the demonstration of our proposed framework. This study can be used in future research endeavors that can further deepen our understanding of loneliness and its societal implications.

Conflicts of Interest

None declared.

References

1. Hards E, Loades ME, Higson-Sweeney N, Shafran R, Serafimova T, Brigden A, et al. Loneliness and mental health in children and adolescents with pre-existing mental health problems: A rapid systematic review. Br J Clin Psychol 2022 Jun;61(2):313-334. [doi: 10.1111/bjc.12331] [Medline: 34529837]
2. Shah SG, Nogueras D, van Woerden HC, Kiparoglou V. Evaluation of the effectiveness of digital technology interventions to reduce loneliness in older adults: systematic review and meta-analysis. J Med Internet Res 2021 Jun 04;23(6):e24712 [FREE Full text] [doi: 10.2196/24712] [Medline: 34085942]
3. Al-Garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: a systematic review. J Biomed Inform 2016 Aug;62:1-11 [FREE Full text] [doi: 10.1016/j.jbi.2016.05.005] [Medline: 27224846]
4. Peytrignet S, Garforth-Bles S, Keohane K. Loneliness monetisation report: analysis for the Department for Digital, Culture, Media and Sport. GOV.UK. 2020 Jun 17. URL: https://assets.publishing.service.gov.uk/media/602fcb91d3bf7f72154fabc3/Loneliness_monetisation_report_V2.pdf [accessed 2025-08-22]
5. Ryall J. Japan: 'Minister of loneliness' tackles mental health crisis. DW. 2021 Apr 23. URL: https://www.dw.com/en/japan-minister-of-loneliness-tackles-mental-health-crisis/a-57311880 [accessed 2025-12-09]
6. Jia Q, Guo Y, Wang G, Barnes SJ. Big data analytics in the fight against major public health incidents (including COVID-19): a conceptual framework. Int J Environ Res Public Health 2020 Aug 25;17(17):6161 [FREE Full text] [doi: 10.3390/ijerph17176161] [Medline: 32854265]
7. 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]
8. Yu Z, Yang X, Dang C, Wu S, Adekkanattu P, Pathak J, et al. A study of social and behavioral determinants of health in lung cancer patients using transformers-based natural language processing models. AMIA Annu Symp Proc 2021;2021:1225-1233 [FREE Full text] [Medline: 35309014]
9. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc Int AAAI Conf Web Soc Media 2014 May 16;8(1):216-225. [doi: 10.1609/icwsm.v8i1.14550]
10. Russell D, Peplau LA, Ferguson ML. Developing a measure of loneliness. J Pers Assess 1978 Jun;42(3):290-294. [doi: 10.1207/s15327752jpa4203_11] [Medline: 660402]
11. Wongpakaran N, Wongpakaran T, Pinyopornpanish M, Simcharoen S, Suradom C, Varnado P, et al. Development and validation of a 6-item Revised UCLA Loneliness Scale (RULS-6) using Rasch analysis. Br J Health Psychol 2020 May 30;25(2):233-256. [doi: 10.1111/bjhp.12404] [Medline: 31999891]
12. Hudiyana J, Lincoln TM, Hartanto S, Shadiqi MA, Milla MN, Muluk H, et al. How universal is a construct of loneliness? Measurement invariance of the UCLA Loneliness Scale in Indonesia, Germany, and the United States. Assessment 2022 Dec 23;29(8):1795-1805. [doi: 10.1177/10731911211034564] [Medline: 34301150]
13. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]

XSL•FO
RenderX

14. Shah HA, Househ M. Mapping loneliness through social intelligence analysis: a step towards creating global loneliness map. BMJ Health Care Inform 2023 Oct 12;30(1):e100728 [FREE Full text] [doi: 10.1136/bmjhci-2022-100728] [Medline: 37827723]
15. Shah HA, Househ M. Mapping loneliness through comparative analysis of USA and India using social intelligence analysis. BMC Public Health 2024 Jan 22;24:253. [doi: 10.1186/s12889-023-17630-3]
16. Shah HA, Househ M. Understanding loneliness through analysis of Twitter and Reddit data: comparative study. Interact J Med Res 2025 Mar 14;14:e49464. [doi: 10.2196/49464] [Medline: 40085832]
17. Jagfeld G, Lobban F, Rayson P, Jones SH. Understanding who uses Reddit: profiling individuals with a self-reported bipolar disorder diagnosis. ArXiv Preprint posted online on April 23, 2021 [FREE Full text] [doi: 10.18653/v1/2021.clpsych-1.1]

Abbreviations

API: application programming interface
NLTK: Natural Language Toolkit
UCLA: University of California, Los Angeles
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by E Mensah; submitted 24.Apr.2024; peer-reviewed by LI Barbosa-Santillan, C Hudak; comments to author 29.Jun.2024; revised version received 17.Sep.2024; accepted 03.Oct.2024; published 15.Jan.2026.

Please cite as:
Shah HA, Househ M, Alsumait L, Alfarhan A
Demonstrating a Social Intelligence Analysis Framework for Loneliness: Infodemiology Approach
Online J Public Health Inform 2026;18:e59861
URL: https://ojphi.jmir.org/2026/1/e59861
doi:10.2196/59861
PMID:

©Hurmat Ali Shah, Mowafa Househ, Loulwah Alsumait, Altaf Alfarhan. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 15.Jan.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Association Between Access to Health Information and Frailty in Older Japanese Adults: Web-Based Cross-Sectional Study

Noriko Hori[1*], MPH, DrPH; Jiaqi Li[1,2*], MPH, PhD; Yosuke Osuka[1*], PhD

[1]Department of Frailty Research, Research Institute, Center for Gerontology and Social Science, National Center for Geriatrics and Gerontology, 7-430 Morioka-cho, Obu City, Japan

[2]Department of Public Health, Faculty of Medicine, Kindai University, Sakai, Japan

[*]all authors contributed equally

Corresponding Author:
Noriko Hori, MPH, DrPH
Department of Frailty Research, Research Institute, Center for Gerontology and Social Science, National Center for Geriatrics and Gerontology, 7-430 Morioka-cho, Obu City, Japan

Abstract

Background: Older adults often access traditional media, such as newspapers, magazines, television, and radio, for health information. However, compared with older adults without frailty, older adults with frailty experience greater declines in physical functions and mental health (including depressive symptoms), as well as social functioning, due to reduced interaction with others, which limits their access to these sources of information.

Objective: This study aimed to identify the health information sources that are less accessible to participants with frailty than to those without frailty.

Methods: A cross-sectional web-based survey was conducted among independent Japanese adults aged ≥75 years. We assessed frailty using the Questionnaire of Medical Checkup for Old-Old, with a score of ≥4 indicating frailty. Participants were asked whether they had accessed any health information source in the past year, including medical institutions, family members, friends or acquaintances, neighbors, government agencies, long-term care or welfare services, television, radio, the internet, magazines, newspapers, or books. The primary explanatory variable was frailty status. Covariates included age, sex, income, education, living arrangements, and health literacy, measured using the eHealth Literacy Scale.

Results: In total, 1032 participants (n=518, 50.2% male; median age: 77 y) were analyzed. Multivariable logistic regression analysis revealed that participants with frailty had significantly less access to the following sources of information compared to individuals without frailty: family (odds ratio [OR] 0.69, 95% CI 0.50 - 0.95), friends/acquaintances (OR 0.70, 95% CI 0.51 - 0.98), radio (OR 0.50, 95% CI 0.31 - 0.79), and newspapers (OR 0.66, 95% CI 0.50 - 0.88). Sex-based subgroup analyses revealed no significant interaction effects, indicating no heterogeneity in the findings.

Conclusions: Older adults with frailty were less likely to obtain health information from interpersonal and traditional media sources than did individuals without frailty. Health information providers need to devise strategies for delivering accurate information and improving usability to enable older adults with frailty to proactively access diverse health information.

(Online J Public Health Inform 2026;18:e83642) doi:10.2196/83642

KEYWORDS

health literacy; frailty; internet; web-based survey; health information sources

Introduction

Access to reliable health information is essential for individuals to recognize health concerns and engage in appropriate health management behaviors [1-3]. Among community-dwelling older adults, traditional media such as newspapers, magazines, television, and radio remain the predominant sources of health information, while interpersonal sources also play a meaningful role [3-5]. Greater access to diverse information sources has been associated with better health status and improved problem-solving abilities regarding health issues [6], highlighting the importance of understanding disparities in access among older adult populations.

Frailty is a clinical geriatric condition characterized by diminished physiological reserves and reduced resilience to stressors, contributing to a heightened risk of disability and adverse outcomes [7]. In Japan, the prevalence of frailty among adults aged ≥65 years is estimated at 7.4%, increasing sharply with age: 10% at ages 75 - 79, 20.4% at ages 80 - 84, and 35.1% at ages ≥85 [8]. Although frailty is potentially reversible, timely access to accurate and actionable health information may play a vital role in preventing severe deterioration, facilitating early

XSL•FO
RenderX

access to support services, and maintaining health-related behaviors.

Older adults with frailty often experience reduced physical, cognitive, and social functioning compared with individuals without frailty [9-11], and declines in vision and hearing have also been linked to frailty [12-15]. These limitations may pose substantial barriers to accessing both interpersonal sources and traditional media, potentially restricting opportunities to obtain essential health information. Prior research has indicated that insufficient social support is associated with the progression of frailty and that access to multiple information sources contributes to healthier behaviors [6,16]. However, it remains unclear which specific types of information sources older adults with frailty are less likely to access. Addressing this knowledge gap is crucial for public health informatics, as targeted communication strategies must be grounded in an accurate understanding of how frailty affects information acquisition.

Therefore, this study aimed to compare access to health information sources between community-dwelling older adults with and without frailty and to identify specific sources that may be less accessible among those with frailty. Findings from this study may inform the development of tailored health communication strategies and interventions to reduce information disparities among vulnerable older populations.

Methods

Study Design and Setting

This study was a cross-sectional internet-based survey conducted over 2 days, from June 12 to 13, 2024.

Participants

Participants were recruited through a panel management company affiliated with Cross Marketing Inc, a Japanese internet research firm. Approximately 50,000 individuals aged ≥75 years were registered as potential respondents to this survey. This study was designed to collect data from 1000 male and female participants across Japan, aged ≥75 years, who were certified as not requiring long-term care. Notably, 10,807 panel members aged ≥75 years were identified as potential respondents. Each panelist could log onto a secure website using a unique ID and password. Recruitment was conducted through this protected site, and interested individuals were asked to read and complete an online informed consent form. Only those who agreed to participate were allowed to complete the questionnaire. Responses were recorded only if all survey items were completed.

Eligibility Criteria

Inclusion criteria were as follows: (1) age ≥75 years, (2) not certified as requiring support or long-term care under the national long-term care insurance system, and (3) provision of informed consent after reviewing the study information. Participant recruitment was terminated when responses exceeded 1000 eligible individuals.

Outcome Measures

Based on previous studies [17-19], the primary outcomes were defined as the use or nonuse (0=not used, 1=used) of 12 health information sources over the past year: health care providers, family members, friends/acquaintances, neighbors, government agencies, long-term care and welfare professionals, television, radio, internet, newspapers, magazines, and books.

Frailty Assessment

Frailty was assessed using the Questionnaire of Medical Checkup for Old-Old (QMCOO) [20], a 15-item screening tool that evaluates domains such as physical function, physical activity and falls, cognitive function, and social participation. Since 2020, the QMCOO has been widely implemented as part of health checkup questionnaires for individuals aged ≥75 years across municipalities in Japan. Moreover, it has demonstrated construct validity for frailty assessment [20], criterion-related validity [21], and predictive validity for mortality and new long-term care certification [22]. Table S1 in Multimedia Appendix 1 lists the items and response options for the QMCOO. Responses indicating poorer health or function were scored 1 point, whereas favorable responses were scored 0 points. These answers are used to calculate points for each question, and the scores range from 0 to 15. A total score of ≥4 points was classified as frailty [21,23]. For the analysis, frailty was treated as a binary variable (0=nonfrail, 1=frail).

Covariates

Covariates were selected with reference to prior studies and included age, sex, final educational attainment level, household income, marital status, the presence of children, living alone [4,19,24], and health literacy. The final educational attainment level was classified into four categories: up to high school, technical or junior college, university or graduate school, and other. Household income was divided into six categories [25]: <2 million yen, 2 to <3 million yen, 3 to <4 million yen, 4 to <5 million yen, ≥5 million yen, and “don’t know/prefer not to answer” (1 million yen=US $6460).

Health literacy was measured using the 8-item eHealth Literacy Scale (eHEALS) [26], presented in Table S2 in Multimedia Appendix 1. Each item was rated on a 5-point Likert scale (1=strongly disagree to 5=strongly agree; total score range: 8 - 40). Following prior research, total scores were dichotomized at the median (0=high, 1=low).

Statistical Analysis

Categorical and continuous variables were summarized as percentages and medians with interquartile ranges, respectively. To examine the association between frailty and each health information source, a multivariable logistic regression analysis was conducted.

After confirming the absence of multicollinearity, the models were specified as follows.

The dependent variable was the use (0) or nonuse (1) of each source, the explanatory variable was frailty, and the covariates included health literacy and demographic variables. All covariates were entered simultaneously into the model.

XSL•FO
RenderX

To assess the robustness of the findings, subgroup analyses based on sex were performed. Subsequently, the interaction terms were tested for statistical significance.

All analyses excluded individuals who discontinued the survey or had missing responses. The survey company also automatically excluded fraudulent responses based on the following criteria: (1) extremely short response times suggesting that items were not read and (2) identical response options recorded for all items.

All analyses were performed using SPSS (version 29; IBM Japan) with a two-sided significance level of 5%.

Ethical Considerations

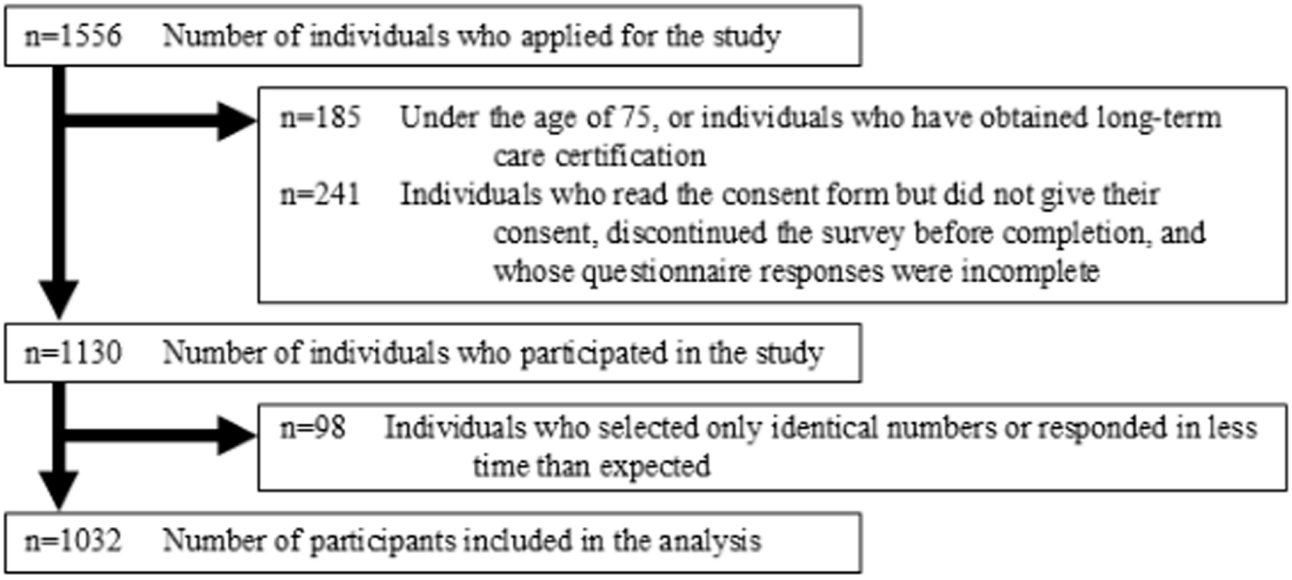
This study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Conflict of Interest and Ethics Committee of the National Center for Geriatrics and Gerontology (approval number 1800). Informed consent was obtained from all participants via a web-based platform. The survey company provided deidentified data. As compensation, participants received points equivalent to several tens to several hundred Japanese yen, according to the survey company's incentive policy.

Results

Participant Enrollment and Characteristics

The participant flow is shown in Figure 1. Among the 1130 individuals who participated in the survey, 98 (9%) were excluded due to extremely short response times or uniform responses across all items, resulting in a final analytical sample of 1032 (1032/1130, 91.3%) participants. Table 1 presents the characteristics of participants stratified by frailty status. Using the QMCOO, 318 (30.8%) participants were classified as frail. Among the 318 participants with frailty, 218 (68.6%) reported having used at least one source of health information. The sources of health information used by older adults with frailty included television (n=173, 54.4%), the internet (n=159, 50%), health care providers (n=147, 46.2%), newspapers (n=112, 35.2%), and family members (n=73, 23%). The median (IQR) eHEALS score was 28 (IQR 24-32) among participants without frailty and 25 (IQR 21-30) among participants with frailty, indicating lower health literacy in the frail group.

Figure 1. Flow of participants.



XSL•FO
RenderX

Table . Participant characteristics.

	All participants (N=1032)	Nonfrail (n=714)	Frail (n=318)
		QMCOO[a] ≤3 points	QMCOO ≥4 points
Age (years), median (IQR)	77 (75-79)	77 (75-79)	77 (76-79)
Sex, male, n (%)	518 (50.2)	339 (47.5)	179 (56.3)
Final educational attainment level, n (%)			
Up to high school	456 (44.2)	302 (42.3)	154 (48.4)
Technical or junior college	161 (15.6)	113 (15.8)	48 (15.1)
University or graduate school	413 (40)	298 (41.7)	115 (36.2)
Other	2 (0.2)	1 (0.1)	1 (0.3)
Household income[b], n (%)			
<2 million yen	131 (12.7)	77 (10.8)	54 (17)
2–<3 million yen	178 (17.2)	116 (16.2)	62 (19.5)
3–<4 million yen	213 (20.6)	143 (20)	70 (22)
4–<5 million yen	120 (11.6)	91 (12.7)	29 (9.1)
≥5 million yen	195 (18.9)	148 (20.7)	47 (14.8)
Don't know/prefer not to answer	195 (18.9)	139 (19.5)	56 (17.6)
Marital status, married	989 (95.8)	687 (96.2)	302 (95)
Presence of children, yes	901 (87.3)	634 (88.8)	267 (84)
Living alone, yes	187 (18.1)	120 (16.8)	67 (21.1)
Number of media sources regularly used, median (IQR)	4 (3-4)	3 (3-4)	4 (3-5)
Television, n (%)	986 (95.5)	692 (96.9)	294 (92.5)
Radio, n (%)	363 (35.2)	268 (37.5)	95 (29.9)
Internet, n (%)	904 (87.6)	634 (88.8)	270 (84.9)
Newspapers, n (%)	716 (69.4)	512 (71.7)	204 (64.2)
Magazines, n (%)	212 (20.5)	161 (22.5)	51 (16)
Books, n (%)	318 (30.8)	240 (33.6)	78 (24.5)
Accessing health information, yes, n (%)	780 (75.6)	562 (78.7)	218 (68.6)
Number of health information sources, median (IQR)	8 (3-10)	8 (4-10)	8 (0-10)
Health information sources, n (%)			
Health care providers	494 (47.9)	347 (48.6)	147 (46.2)
Family members	317 (30.7)	244 (34.2)	73 (23)
Friends/acquaintances	293 (28.4)	225 (31.5)	68 (21.4)
Neighbors	63 (6.1)	49 (6.9)	14 (4.4)
Government agencies	213 (20.6)	163 (22.8)	50 (15.7)
Long-term care and welfare professionals	45 (4.4)	33 (4.6)	12 (3.8)
Television	641 (62.1)	468 (65.5)	173 (54.4)
Radio	134 (13)	109 (15.3)	25 (7.9)
Internet	593 (57.5)	434 (60.8)	159 (50)
Newspapers	464 (45)	352 (49.3)	112 (35.2)
Magazines	100 (9.7)	73 (10.2)	27 (8.5)

XSL•FO
RenderX

	All participants (N=1032)	Nonfrail (n=714)	Frail (n=318)
		QMCOO[a] ≤3 points	QMCOO ≥4 points
Books	129 (12.5)	101 (14.1)	28 (8.8)
QMCOO score, median (IQR)	2 (1-4)	2 (1-2)	5 (4-6)
eHEALS[c] score, median (IQR)	28 (24-31)	28 (24-32)	25 (21-30)

[a]QMCOO: Questionnaire of Medical Checkup for Old-Old.

[b]1 million yen=US $6460.

[c]eHEALS: eHealth Literacy Scale.

Association Between Health Information Sources and Frailty

Table 2 presents the results of multivariable logistic regression analyses examining the association between frailty and nonaccess to each health information source. Compared with participants without frailty, participants with frailty were less likely to access the following sources of health information: family members (OR 0.69, 95% CI 0.50 - 0.95; *P*=.02), friends or acquaintances (OR 0.70, 95% CI 0.51 - 0.98; *P*=.03), radio (OR 0.50, 95% CI 0.31 - 0.79; *P*=.004), and newspapers (OR 0.66, 95% CI 0.50 - 0.88; *P*=.005).

Table . Adjusted multivariable logistic regression analyses of the association between health information sources and frailty. The covariates included are age, sex, final educational attainment level, household income, marital status, the presence of children, living alone, and health literacy.

	Multivariable model, odds ratio (95% CI)	*P* value
Health care providers	1.08 (0.82 - 1.43)	.46
Family members	0.69 (0.50 - 0.95)	.02
Friends/acquaintances	0.70 (0.51 - 0.98)	.03
Neighbors	0.79 (0.42 - 1.49)	.67
Government agencies	0.75 (0.52 - 1.07)	.14
Long-term care and welfare professionals	1.11 (0.55 - 2.28)	.90
Television	0.76 (0.57 - 1.00)	.06
Radio	0.50 (0.31 - 0.79)	.004
Internet	0.83 (0.62 - 1.11)	.46
Newspapers	0.66 (0.50 - 0.88)	.005
Magazines	1.03 (0.63 - 1.69)	.78
Books	0.79 (0.50 - 1.26)	.29

Sex-based subgroup analyses revealed no significant interaction effects for these associations (family: *P*=.47; friends/acquaintances: *P*=.28; radio: *P*=.61; newspapers: *P*=.23), indicating that the associations were consistent across male and female participants.

Discussion

Novelty of the Study

To the best of our knowledge, this study is the first to examine the relationship between frailty and sources of health information among adults aged ≥75 years. Although previous studies have identified commonly accessed sources of health information among the general older population [3-5], studies focusing specifically on older adults with frailty are limited. Our findings revealed that older adults with frailty were significantly less likely than individuals without frailty to access family members, friends/acquaintances, newspapers, and radio as information sources. In addition, these associations did not differ by sex. These findings are important for understanding the extent to which older adults with frailty have difficulty accessing certain health information sources, and they provide fundamental evidence for examining methods and support strategies for delivering health information tailored to individuals with frailty by clarifying which communication channels are more susceptible to vulnerability. In particular, reduced use of interpersonal information sources, such as family and friends, may reflect weakened social connections and suggests priority issues that should be addressed in efforts to reduce information disparities and strengthen long-term care prevention policies.

Relationship Between Health Information Sources and Frailty

Family and Friends/Acquaintances

Recent conceptualizations of frailty have expanded beyond physical aspects to include social dimensions. Limited interaction with family, friends, and neighbors, as well as reduced social support, are recognized components of frailty [9-11]. Older adults experiencing social isolation tend to be frail

XSL•FO
RenderX

[27], and those with low social support have an elevated risk of transitioning to frailty [16]. The less frequent use of family and friends as health information sources among individuals with frailty observed in this study might reflect insufficient social support, possibly associated with frailty status.

Newspapers

Low health literacy has been associated with frailty [28,29], and individuals with low socioeconomic status or educational attainment tend to have a reduced ability to understand printed materials such as newspapers [30]. In this study, participants with frailty had lower eHEALS scores, household income, and final educational attainment level than individuals without frailty. Notably, the average income of older households is approximately 3 million yen (US $19,390) [24]. Our findings revealed that approximately 37% of the participants with frailty (compared with 27% of those who were not frail) are in this category. As this study was conducted online, some participants with frailty possibly accessed free news content online or refrained from subscribing to print newspapers for economic reasons. Moreover, declines in visual acuity are associated with frailty [13], and access to newspapers printed in small fonts may have been limited.

Radio

In rural areas of Japan, many older adults access multiple media sources such as television, radio, and newspapers, which positively influence health behaviors [6]. However, studies abroad suggest that adults aged ≥85 years tend to trust the radio less as a health information source than those aged ≤84 years. Further, listening to the radio presents two challenges: (1) the inability to ask questions to clarify information and (2) the criteria for broadcasting content are unclear, making it challenging to evaluate health information [5]. Furthermore, age-related hearing loss, which is commonly associated with frailty [31], may reduce the accessibility of audio-only media such as radio for older adults with hearing loss.

In studies targeting the general adult population and patients with cancer, sources of health information were reported as follows: television and radio (31.9%), newspapers (23.7%), the internet (54% - 56%), physicians (9% - 13%), and books (7% - 12%) [6,30]. Among these, trusted health information sources were most commonly physicians (20.4% - 53%) and health care professionals (12.2% - 60%) [17,18] suggesting that information provided by experts with medical knowledge is considered the most reliable. In this study, the sources of health information among older adults with frailty were analyzed; however, their most trusted sources were not identified. A substantial proportion of older adults with frailty in this study reported obtaining information from health care professionals. Further research is needed to identify more effective sources of health information, including those considered trustworthy.

Strengths

A key strength of this study is its large sample size, comprising >1000 community-dwelling older adults aged ≥75 years recruited across Japan, which provides high statistical power. Moreover, sex-based subgroup analyses revealed no significant interaction effects, supporting the robustness of the main results. By targeting individuals with a higher prevalence of frailty aged ≥75 years, we successfully identified specific health information sources that older adults with frailty are less likely to access. These findings underscore the need to develop more accessible, trustworthy, and effective health information resources tailored to the needs of this population. For instance, scientifically grounded platforms operated by public or medical institutions, community-based online resources, and multimodal delivery formats combining text and audio content should be considered.

Limitations

This study has some limitations. First, as a cross-sectional study, causal inference cannot be established. Second, participants were recruited through a web-based panel. Since older adults with frailty without internet access were likely excluded, sampling bias may have occurred, restricting the generalizability of the findings to relatively healthier, internet-using older adults in Japan. Consequently, the results of this study may underestimate disparities in access to health information among older adults with frailty. In contrast, although the prevalence of frailty among older adults in Japan is reported to be 7.4% [8], the proportion of older adults with frailty identified in this study was 30.8%. This suggests that many older adults who are often excluded from conventional surveys—such as those with difficulty walking, those who tend to remain indoors, and those with limited social interaction and an increased risk of frailty—may have been included in this study. Third, frailty was not assessed using objective measures such as gait speed or grip strength, which may have reduced the accuracy of the assessment. Finally, confounding factors related to health status and cognitive function, such as the frequency and content of access to health information, severe cognitive impairment, major comorbidities, physical functional limitations, and mobility, were not adjusted for.

Conclusions

Our findings suggest that older adults with frailty were less likely to obtain health information from interpersonal sources such as family and friends, as well as from traditional media such as radio and newspapers, than individuals without frailty. These findings are valuable for developing effective and personalized strategies to deliver health information to older adults with frailty.

Acknowledgments

We would like to express our deepest gratitude to all those involved in this study.

Generative artificial intelligence use: We used ChatGPT during the translation process.

XSL•FO
RenderX

Funding

This study was supported by a grant from the Japan Society for the Promotion of Science (grant number 25K20834).

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: NH
Data curation: NH
Formal analysis: NH
Funding acquisition: NH
Methodology: NH, JL, YO
Software: NH
Writing–original draft: NH, JL, YO
Data interpretation, critical revision of the manuscript for important intellectual content, and final confirmation of the manuscript: all authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
Items and response options of the Questionnaire for Medical Checkup of Old-Old and eHealth Literacy Scale.
[DOCX File, 27 KB - ojphi_v18i1e83642_app1.docx]

References

1. Synnot AJ, Hill SJ, Garner KA, et al. Online health information seeking: how people with multiple sclerosis find, assess and integrate treatment information to manage their health. Health Expect 2016 Jun;19(3):727-737. [doi: 10.1111/hex.12253] [Medline: 25165024]
2. Kim SC, Shah DV, Namkoong K, McTavish FM, Gustafson DH. Predictors of online health information seeking among women with breast cancer: the role of social support perception and emotional well-being. J Comput Mediat Commun 2013 Jan;18(2):98-118. [doi: 10.1111/jcc4.12002] [Medline: 24634575]
3. Islam NS, Patel S, Wyatt LC, et al. Sources of health information among select Asian American immigrant groups in New York City. Health Commun 2016;31(2):207-216. [doi: 10.1080/10410236.2014.944332] [Medline: 26266574]
4. Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ. Trends and predictors of COVID-19 information sources and their relationship with knowledge and beliefs related to the pandemic: nationwide cross-sectional study. JMIR Public Health Surveill 2020 Oct 8;6(4):e21071. [doi: 10.2196/21071] [Medline: 32936775]
5. Le T, Chaudhuri S, White C, Thompson H, Demiris G. Trust in health information sources differs between young/middle and oldest old. Am J Health Promot 2014;28(4):239-241. [doi: 10.4278/ajhp.120921-ARB-459] [Medline: 23941098]
6. Morishita-Suzuki K, Watanabe S. Health information source patterns and dietary variety among older adults living in rural Japan. Int J Environ Res Public Health 2024 Jul 1;21(7):865. [doi: 10.3390/ijerph21070865] [Medline: 39063442]
7. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
8. Kojima G, Iliffe S, Taniguchi Y, Shimada H, Rakugi H, Walters K. Prevalence of frailty in Japan: a systematic review and meta-analysis. J Epidemiol 2017 Aug;27(8):347-353. [doi: 10.1016/j.je.2016.09.008] [Medline: 28142044]
9. Gobbens RJJ, van Assen M, Luijkx KG, Wijnen-Sponselee MT, Schols J. The Tilburg Frailty Indicator: psychometric properties. J Am Med Dir Assoc 2010 Jun;11(5):344-355. [doi: 10.1016/j.jamda.2009.11.003] [Medline: 20511102]
10. Garre-Olmo J, Calvó-Perxas L, López-Pousa S, de Gracia Blanco M, Vilalta-Franch J. Prevalence of frailty phenotypes and risk of mortality in a community-dwelling elderly cohort. Age Ageing 2013 Jan;42(1):46-51. [doi: 10.1093/ageing/afs047] [Medline: 22454134]
11. Teo N, Gao Q, Nyunt MSZ, Wee SL, Ng TP. Social frailty and functional disability: findings from the Singapore longitudinal ageing studies. J Am Med Dir Assoc 2017 Jul 1;18(7):637. [doi: 10.1016/j.jamda.2017.04.015] [Medline: 28648903]
12. Kawaguchi K, Abe N, Hiratsuka Y, Kojima K, Kondo K. Self-reported hearing and vision impairment and incident frailty in Japanese older people: a 3-year longitudinal analysis of the Japan Gerontological Evaluation Study. Arch Gerontol Geriatr 2023 Jan;104:104834. [doi: 10.1016/j.archger.2022.104834] [Medline: 36257161]
13. Hou T, Liu M, Zhang J. Bidirectional association between visual impairment and frailty among community-dwelling older adults: a longitudinal study. BMC Geriatr 2022 Aug 15;22(1):672. [doi: 10.1186/s12877-022-03365-0] [Medline: 35971062]

XSL•FO
RenderX

14. Tian R, Almeida OP, Jayakody DMP, Ford AH. Association between hearing loss and frailty: a systematic review and meta-analysis. BMC Geriatr 2021 May 25;21(1):333. [doi: 10.1186/s12877-021-02274-y] [Medline: 34034656]
15. Assi S, Zhang W, Carey JP, et al. The association of hearing loss with frailty among community-dwelling older adults: findings from the National Health and Aging Trends Study. BMC Geriatr 2023 Nov 17;23(1):754. [doi: 10.1186/s12877-023-04465-1] [Medline: 37978436]
16. Barghouth MH, Klein J, Bothe T, Ebert N, Schaeffner E, Mielke N. Social support and frailty progression in community-dwelling older adults. Front Public Health 2024;12:1408641. [doi: 10.3389/fpubh.2024.1408641] [Medline: 39086799]
17. A nationally representative cross‐sectional survey on health information access for consumers in Japan 2023 (INFORM study 2023). National Cancer Center, Institute for Cancer Control and Prevention. URL: https://www.ncc.go.jp/jp/icc/behav-sci/project/070/index.html [accessed 2025-12-02]
18. Health awareness survey report [Report in Japanese]. : Ministry of Health, Labour and Welfare; 2018 URL: https://www.mhlw.go.jp/file/04-Houdouhappyou-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/002.pdf [accessed 2025-12-02]
19. Mitsutake S, Takahashi Y, Otsuki A, et al. Chronic diseases and sociodemographic characteristics associated with online health information seeking and using social networking sites: nationally representative cross-sectional survey in Japan. J Med Internet Res 2023 Mar 2;25:e44741. [doi: 10.2196/44741] [Medline: 36862482]
20. Ishizaki T, Masui Y, Nakagawa T, et al. Construct validity of a new health assessment questionnaire for the national screening program of older adults in Japan: the SONIC study. Int J Environ Res Public Health 2022 Aug 19;19(16):10330. [doi: 10.3390/ijerph191610330] [Medline: 36011962]
21. Hori N, Ishizaki T, Masui Y, et al. Criterion validity of the health assessment questionnaire for the national screening program for older adults in Japan: the SONIC study. Geriatr Gerontol Int 2023 Jun;23(6):437-443. [doi: 10.1111/ggi.14585] [Medline: 37096927]
22. Hori N, Li J, Kinoshita K, Yoshiura K, Osuka Y, Satake S. Predictive validity of the Questionnaire for Medical Checkup of Old-Old for all-cause mortality and disability incidence. Geriatr Gerontol Int 2024 Nov;24(11):1203-1209. [doi: 10.1111/ggi.14977] [Medline: 39317380]
23. Li J, Yasuoka M, Kinoshita K, et al. Validity of the Questionnaire for Medical Checkup of Old-Old (QMCOO) in screening for physical frailty in Japanese older outpatients. Geriatr Gerontol Int 2022 Oct;22(10):902-903. [doi: 10.1111/ggi.14472] [Medline: 36043845]
24. Yamaguchi K, Akahane T, Yasuda E, Akahane M. Age-specific differences in association between personality and changes in outing behaviors during the COVID-19 pandemic in Japan: cross-sectional web-based questionnaire survey. Online J Public Health Inform 2025 Jun 12;17:e63120. [doi: 10.2196/63120] [Medline: 40505243]
25. Overview of the average results for 2024: national survey of family income and expenditure p 17 [Report in Japanese]. : Ministry of Internal Affairs and Communications URL: https://www.stat.go.jp/data/kakei/sokuhou/tsuki/pdf/fies_gaikyo2024.pdf [accessed 2026-02-18]
26. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. Developing Japanese version of the eHealth Literacy Scale (eHEALS). Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
27. Kojima G, Aoyama R, Tanabe M. Associations between social isolation and physical frailty in older adults: a systematic review and meta-analysis. J Am Med Dir Assoc 2022 Nov;23(11):e3-e6. [doi: 10.1016/j.jamda.2022.06.012] [Medline: 35810790]
28. Huang CH, Lai YC, Lee YC, Teong XT, Kuzuya M, Kuo KM. Impact of health literacy on frailty among community-dwelling seniors. J Clin Med 2018 Nov 26;7(12):481. [doi: 10.3390/jcm7120481] [Medline: 30486284]
29. Shirooka H, Nishiguchi S, Fukutani N, et al. Association between comprehensive health literacy and frailty level in community-dwelling older adults: A cross-sectional study in Japan. Geriatr Gerontol Int 2017 May;17(5):804-809. [doi: 10.1111/ggi.12793] [Medline: 27381868]
30. Cutilli CC, Simko LC, Colbert AM, Bennett IM. Health literacy, health disparities, and sources of health information in U.S. older adults. Orthop Nurs 2018;37(1):54-65. [doi: 10.1097/NOR.0000000000000418] [Medline: 29369135]
31. Kamil RJ, Betz J, Powers BB, et al. Association of hearing impairment with incident frailty and falls in older adults. J Aging Health 2016 Jun;28(4):644-660. [doi: 10.1177/0898264315608730] [Medline: 26438083]

Abbreviations

eHEALS: eHealth Literacy Scale
OR: odds ratio
QMCOO: Questionnaire for Medical Checkup of Old-Old

XSL•FO
RenderX

Edited by E Mensah; submitted 06.Sep.2025; peer-reviewed by H Siamian, LIN Chen; accepted 29.Dec.2025; published 27.Feb.2026.

Please cite as:
Hori N, Li J, Osuka Y
Association Between Access to Health Information and Frailty in Older Japanese Adults: Web-Based Cross-Sectional Study
Online J Public Health Inform 2026;18:e83642
URL: https://ojphi.jmir.org/2026/1/e83642
doi:10.2196/83642

© Noriko Hori, Jiaqi Li, Yosuke Osuka. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 27.Feb.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigating Strategies for Sustaining Modern Health Information Systems for Enhanced Care Delivery in Fragile Resource-Constrained Settings Through Electronic Medical Record Implementation in Haiti: Qualitative Study

Benito D Isaac[1*], MMSc; Annie Michaelis[2*], PhD; Meredith Casella Jean-Baptiste[3*], CNM; Loune G Viaud[2,3], BA; Wesler Lambert[3], MD

[1]Department of Global Health and Social Medicine, Harvard Medical School, Harvard University, 641 Huntington Avenue, Boston, MA, United States
[2]Partners In Health (PIH), Boston, MA, United States
[3]Zanmi Lasante (ZL), Port-au-Prince, Haiti
[*]these authors contributed equally

Corresponding Author:
Benito D Isaac, MMSc
Department of Global Health and Social Medicine, Harvard Medical School, Harvard University, 641 Huntington Avenue, Boston, MA, United States

Abstract

Background: Robust and reliable health information systems (HISs) are foundational to equitable health care delivery in resource-constrained settings. Yet, HISs often exhibit significant fragmentation and complexity, which stem from many factors, including inadequate infrastructure, limited and unevenly allocated financial resources, expertise gaps, and a lack of integrated systems. At the same time, advances in modern HISs and digital technologies, such as electronic medical records (EMRs), present opportunities for addressing these limitations and supporting evidence-based health systems if well implemented and sustained. However, limited attention has been paid to how modern and resilient HISs can be effectively sustained in fragile, resource-constrained settings.

Objective: This empirical study seeks to identify pathways for strengthening the resilience, adaptability, and contextual-fit of EMRs in fragile, resource-constrained settings such as Haiti.

Methods: Using a qualitative research methodology, the study used semistructured interviews with purposive sampling of key informants, including frontline doctors, nurses, and IT or data specialists. Interview participants were selected for their expertise and capacity to offer insightful and varied viewpoints on the topic. Transcripts were analyzed using an inductive approach to identify key emerging themes.

Results: The findings of this investigation reveal that implementing resilient EMR in limited-resource settings, such as Haiti, requires a comprehensive strategy that accounts for ecosystemic challenges from interdependent systems (electricity, sociopolitical instability, internet connectivity), as well as the intrinsic complexities of legacy systems.

Conclusions: Drawing on empirical evidence, the study identifies a set of protective strategies that, if effectively implemented, may enhance the resilience and adaptability of EMRs. These strategies include prioritizing integration and interoperability between systems, building redundancy to mitigate cascading failures derived from other interdependent systems, strengthening staffing to support system use, right-sizing the paper footprint (eg, improving handwriting-to-text scanning and digitization solutions), and embracing technological innovations.

(Online J Public Health Inform 2026;18:e82396) doi:10.2196/82396

KEYWORDS

digital health; electronic medical records; Haiti; low- and middle-income country; LMIC; health information systems; OpenMRS; qualitative methods

Introduction

Robust and reliable health information systems (HISs) are foundational to equitable health care delivery in resource-constrained settings [1]. High-functioning HISs enable timely, evidence-based decision-making across all levels of the health system—from clinical decision support to broader program and policy management—while improving workflows and resource allocations [2]. Yet, HISs in such settings often

XSL•FO
RenderX

exhibit considerable fragmentation, complexity, and lack of interoperability, stemming from factors that include, but are not limited to, inadequate infrastructure [3], limited financing [4-7], expertise gaps, lack of integrated enterprise systems, and uneven endowment of resources across geographical areas [8,9]. These constraints hinder the ability of HISs to effectively support health care delivery by undermining data quality, disrupting workflows, and weakening continuity of care with consequences for patient outcomes, service efficiency, and decision-making [9,10].

Concurrently, advances in modern HISs and digital technologies, such as electronic medical records (EMRs), present opportunities for addressing these limitations and supporting evidence-based health systems if well implemented and sustained. Existing literature on the topic has primarily focused on describing the main characteristics of HISs in low- and middle-income countries (LMICs) and identifying common barriers and enablers of EMR implementation in such contexts. Yet, less attention has been paid to how EMR systems evolve and are sustained over time in fragile health systems where successful implementation is shaped not only by technical design but also by political and social instability, competing stakeholders' priorities, chronic resource constraints, and the persistent gap between the promise of modern HISs and their practical realization.

Haiti provides a particularly important case through which to understand these dynamics. In the early 2000s, the introduction of an EMR platform for health care data management in a Haitian public hospital, though limited in scope, marked a significant milestone in the country's digital health development. Initially launched as part of the multidrug-resistant tuberculosis and HIV program, this pioneering initiative improved care for patients with HIV in the country [11] and contributed to the later development of OpenMRS [12], an open-source EMR platform widely adopted by governments, nonprofit organizations, and other health care institutions globally. Therefore, this experience offers a concrete and valuable opportunity to understand how modern platforms are introduced, adapted, and maintained in fragile, resource-constrained settings.

Drawing upon sociotechnical theoretical frameworks and empirical evidence, this study seeks to address this critical gap by integrating the perspectives of data or IT experts and frontline health care providers to identify pathways for strengthening the resilience, adaptability, and contextual-fit of EMRs in complex, resource-constrained settings such as Haiti.

Methods

Study Design

This study used a qualitative research design centered on in-depth, semistructured interviews with key informants, selected through purposive sampling based on their relevant expertise, professional proximity to EMR implementation, and presumed ability to provide well-informed and diverse perspectives on the subject matter.

Study Setting

The study was conducted within 2 public health care facilities operated by the Haitian Ministry of Health (MSPP) and supported by the Haitian health care nonprofit organization, Zanmi Lasante (ZL; the sister organization of the global health nonprofit, Partners In Health). The 2 facilities were selected as information-rich cases that capture meaningful variation in resource endowment, thereby improving the transferability of the findings to comparable LMIC settings. One of these facilities is a tertiary referral teaching hospital, Hôpital Universitaire de Mirebalais (HUM), and the other is a primary health care center, Saint-Michel Health Center of Boucan-Carré. These health care facilities are located in the communes of Mirebalais and Boucan-Carré, respectively, both within the rural Central Plateau department of Haiti.

Sampling and Recruitment

The researchers used a purposive sampling method to select study participants based on job function, tenure at ZL or Partners In Health, and location of work to ensure diversity across geographic locations and relevant professional backgrounds. An equal number of health care providers were chosen from both HUM and Boucan Carré study sites. Information systems specialists with extensive experience working for the institution both in the Partners In Health Boston office and at HUM, Boucan Carré, and the ZL central office in Haiti were also included, thereby providing a varied range of perspectives and expertise.

Data Collection

An in-depth interview guide was developed based on study objectives and relevant literature. It included open-ended questions with the flexibility to elicit participants' experiences and reflections about the topic. The guide was tested and adapted during the design process. The participants in the test phase were excluded from the final data collection. Key informant interviews were completed in-person or through virtual video conference meetings. To ensure confidentiality, both in-person and virtual interviews were conducted in private rooms. The duration of interviews varied, ranging from approximately 45 minutes to 1 hour and 30 minutes. Interviews were conducted in English or Haitian Creole, contingent upon the native language and preference of each participant. Interviews were conducted until thematic saturation was achieved, defined as the point at which additional interviews no longer generated new themes or substantially enriched analytical perspectives.

Analysis

A research assistant, trained by the principal investigator (PI) in ethical standards and research confidentiality, transcribed the Haitian Creole interviews. The English interviews were conducted remotely via the Zoom app with the transcription feature activated; the transcripts were later reviewed and edited by the PI for accuracy and spelling errors. The PI, who is fluent in English and a native speaker of Haitian Creole, completed all translation tasks from Haitian Creole into English for this research. Prior to analysis, all transcripts were carefully reviewed to ensure accurate translation and fidelity to the original interview audio recordings. The PI also reviewed all

XSL•FO
RenderX

automatically generated Zoom transcripts and corrected any inaccuracies.

Using an inductive, thematic approach to content analysis, the interview transcripts were examined to identify emergent themes relevant to the research study. Subsequently, open coding was performed on 4 transcripts to establish preliminary categories for analysis, providing a foundation for the extraction of key themes. The remaining transcripts were then coded through an iterative process that allowed for continuous reexamination and deepening of codes and emergent themes. We used Dedoose (version 9.05; cloud app for managing, analyzing, and presenting qualitative and mixed method research data) for the data analysis.

Given the PI's insider familiarity with the context, reflexive memo-writing and attention to discrepant cases were used to limit the influence of prior assumptions.

Ethical Considerations

This study obtained institutional review board (IRB) approval from both Harvard University IRB (CR22-0157-01) in the United States and ZL IRB (ZLIRB06092022) in Haiti. Verbal consent was obtained from all interview participants. Names and any other identifiable information were redacted from the transcripts to ensure anonymity. All participants in the interviews were adults, and none were classified as members of vulnerable populations.

Results

Key Informant Characteristics

In this investigation, a total of 18 key informants agreed to participate in individual interviews. One transcript was removed from the analysis due to issues related to audio and transcription quality. As a result, 17 interviews were included in this analysis (Table 1).

Table . Number of key informants, by area of expertise and location.

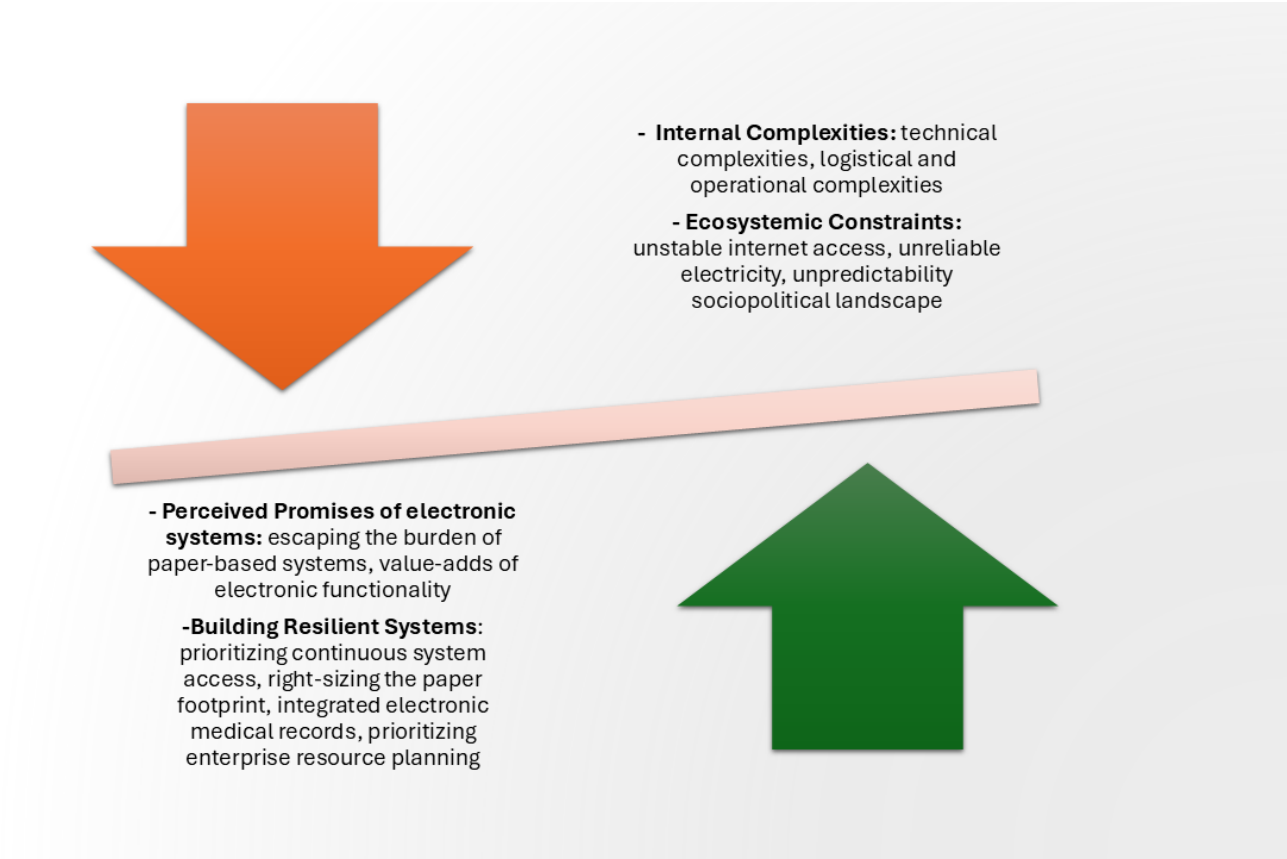
Study site	Health care providers, n (%)	Information system experts, n (%)	Total, n (%)
Hôpital Universitaire Mirebalais	4 (40)	6 (60)	10 (100)
Boucan Carré Health Center	4 (100)	0 (0)	4 (100)
Partners In Health Boston office	0 (0)	3 (100)	3 (100)
Total	8 (47)	9 (53)	17 (100)

Themes

The study findings reveal four primary themes: (1) Perceived promise of electronic systems, (2) Ecosystemic constraints, (3) Internal complexities, and (4) Building resilient information systems (Figure 1).

XSL•FO
RenderX

Figure 1. Illustration of the themes derived from the analysis of interview transcripts.



Theme 1: Perceived Promise of Electronic Systems

Escaping the Burdens of Paper-Based Systems

Key informants shared a range of perspectives on the current and potential future value of transitioning from paper-based records to EMRs. The most frequently cited reason for investing in the transition to electronic systems was related to the perceived cumbersome and unsustainable nature of paper-based data collection. Respondents noted challenges in finding suitable storage space and maintaining proper management of paper files, resulting in negative impacts on patient care. For example:

> *I remember once I went to a site and when I got there, I saw a consultation [room] that was supposed to close because the archive was already full and they started putting files in it. And I said to myself, "this is serious, it's terrible, we have to come up with a plan for paperless."* [IT or data specialist 5]

In addition to the growing physical burden of storing paper records, key informants often described issues related to patient file loss, lack of privacy in handling patient information, untimely archiving of patient records, and the duplication of patient files. One provider clearly explained the health impacts of such losses:

> *Okay, if a patient has a lost file, then it has a significant impact because from the first contact you have with the patient, there is a lot of information that you take from when the child is just born to when you see them at the clinic. This information is important because there is information you have since birth. The child may develop a pathology after that you need to link with the birth history.* [Provider 1]

"Value-Adds" of Electronic Functionality

In addition to describing the ways in which electronic systems can help avoid the many pitfalls of paper-based systems, the key informants also noted a variety of "value-adds" in electronic systems. For example, respondents noted that information in electronic formats is easier to make available when and where it is needed, allowing for more effective data-driven decision-making:

> *Data that is reliable, usable, and can allow for decision-making to improve the quality of care and analyze disparities to see who is benefiting from care, who is missing appropriate care, and how the institution can better provide better coverage and quality of care.* [Provider 4]

Theme 2: Ecosystemic Constraints

Overview

Ecosystemic constraints encompass social, political, and economic challenges at a national level that impact the implementation of HISs. These constraints largely exist outside the direct control of a single institution to address.

Unreliable Electricity

Every key informant identified insufficient electrical capacity and/or power outages as significant obstacles to the effective functioning of EMRs. As one participant noted:

XSL•FO
RenderX

> *Well, the national context doesn't allow for a solely electronic system due to the frequent power outages and internet issues at various independent sites. [...]* [IT or data specialist 8.]

In the context of unreliable electricity, participants highlighted another important challenge: a paradox wherein they felt desperate to escape the "trap" of paper-based systems described above but also believed it was impossible to avoid paper at least as a backup because they anticipated frequent electronic system outages due to unreliable power and internet. For example, one key informant noted:

> *I think we will still use paper because we do not always have electricity, the internet often gives problems, so we will always be able to use paper files. This creates a conundrum for system users who then must spend time on double-data entry in both electronic and paper systems.* [Provider 10]

Unstable Internet Access

Along with comments about unreliable electricity, participants also commonly highlighted challenges related to internet access and the inadequate reliability of internet service providers as significant obstacles to implementing reliable electronic systems. They noted that the inability to reliably connect to the internet has direct implications for EMR systems and the degree to which they can rely on state-of-the-art cloud information storage options. As one key respondent noted:

> *I'm always advocating for more cloud EMRs and the local teams are generally coming back and saying, "that's not an option for us, our internet is not reliable enough, not fast enough".* [IT or data specialist 7]

Several health care provider respondents noted that they prefer using the EMR rather than paper, but that internet connectivity (both functioning internet service providers and the stable power necessary to keep internet and EMR platform running) can impede EMR use:

> *We do have access to the internet, but if the connection is lost, it poses a problem, and we have to write everything down on paper. However, for me personally, the EMR system is easier.* [Provider 1]

Unpredictability of the Sociopolitical Landscape

Participants frequently cited social and political instability in the country as a significant barrier to deploying, implementing, and sustaining information systems. For example, one respondent reported:

> *With political instability and natural disasters, and kind of ongoing safety and security concerns all of those things pose a threat to … a unified EMR and all of these things that we're trying to work towards from a health information systems perspective.* [IT or data specialist 8]

Although this instability affects the ability to maintain functioning systems, respondents noted how these same social and political contexts also contribute to an increase in the number of patients seeking care at University Hospital of Mirebalais (due to deteriorating conditions at other health facilities), putting even more pressure on the systems:

> *Initially, we had a hospital that was built to provide patients with specialized care, such as orthopedic, surgical, and trauma care. Now, things have changed with the political problems and the healthcare system, where all patients come to the HUM hospital. Currently, we have almost 700,000 to 800,000 files.* [IT or data specialist 5]

Theme 3: Internal Complexities

Overview

This theme highlights challenges that are specific to the organization implementing an EMR. Key informants described how these internal complexities—both technical and operational—could endanger the ability of the electronic system to provide value to clinicians and their patients. Among the IT or data specialists interviewed, there was a strong sense of responsibility to address the complexities and find a way to make the electronic systems more fully meet the end users' needs. As one respondent described it:

> *When you implement a system and tell the "provider" that I am simplifying your life, you cannot invent a solution that complicates life. If you invent a solution or propose a […] hybrid system, then you are doing "duplication of tasks" and it will result in more work […]. The quality of data will be altered in both systems, and in the end, you won't help.* [IT or data specialist 8]

Technical Complexities

Several of the challenges mentioned by key informants focused on software development and can be considered "technical" complexities. The first of these is related to integration across multiple existing (and often outdated) systems. Participants reported that current HISs have been deployed for a significant period and are deeply entrenched in the health care system. For example:

> *It's like you have a "Stand-alone Application" managing HIV data and you have another "Stand-alone application" managing NCD [non-communicable diseases] data. You have several applications, even in their initial conception, they did not anticipate that they would integrate or connect with other systems. So, when you have these situations, it becomes more difficult.* [IT or data specialist 3]

Due to strong adoption by users and support from stakeholders (including funders, public partners, and users), these systems can be challenging to modify, even when they are not perfectly meeting stakeholders' needs or when multiple siloed systems would benefit from connecting with each other.

Another technical complexity participants frequently cited is the challenge of developing a master patient index (MPI) across programs and sites within the entire network of care. Respondents flagged that having an MPI is necessary to allow

XSL•FO
RenderX

for interoperability and/or integration between systems so that unique patients can be accurately identified across different health care facilities and clinical programs. Developing an optimal MPI would require investments in technology as well as collaboration with the government and other health care delivery organizations nationally to ensure that patient records can follow a patient regardless of where they seek health care. Several respondents noted that the technical challenges in creating an MPI are magnified by the lack of a functional and comprehensive national ID program. Because of the lack of a standard national system to identify unique individuals, investments must be made in other methods to verify identity (such as biometrics):

> *One of the challenges at the national level, that's a country-wide challenge, is that there aren't many legal documents to allow you to identify patients or access their services. […] you know that there's a part of the population that doesn't even have a birth certificate, let alone other official identity pieces like a National ID Card or passport.* [IT or data specialist 6]

A final type of technical complexity mentioned by key informants relates to the science and art of "requirements gathering" or the tailoring of a software solution to meet all the needs of its varied users. Participants commonly expressed concerns regarding information systems that fail to consider the needs of users and the workflow of health care operations. Due to gaps in the design process, these systems often do not comprehensively incorporate user input, leading to poor user experiences and a lack of solutions tailored to meet real-world needs. Respondents described a growing awareness that more effort must be spent on understanding all nuances of the clinical workflow so that the HIS matches those processes. As one respondent noted:

> *[If] it is not "user-friendly," it is also because the logic of the provider's trade does not reflect it. [...]. So, when you develop a system, you must have a process that allows you to approach your solution as closely as possible with its work and the usual task flow.* [IT or data specialist 8]

Logistical and Operational Complexities

Even after an electronic system has been fully built, there are logistical and operational complexities that can impede optimal use. One of the most common operational challenges noted by key informants was related to user support. Participants noted that insufficient staffing to provide on-site support during the deployment of information systems is a common obstacle that undermines successful system implementation. When doctors and nurses encounter technical problems while using the system, they may struggle to find immediate assistance, which can lead to delays in care provision and lower uptake of the system:

> *When you have a line of patients waiting for you to provide service, the system must be able to perform efficiently, so that when you click on a button, it doesn't take 15 to 20 seconds. It must be performant enough to free up your line as much as possible.* [Provider 3]

Key informants noted that this challenge of user support is affected both by financial resource constraints and by limitations in the ability to find and retain technical staff willing to work at the point of care in the many rural facilities supported by ZL or Partners In Health:

> *You need the money but it's a little bit easier to hire software developers who can work remotely. You know whether it's in Haiti, remotely, or in Boston to develop software. But it's hard, it's a little bit harder to hire people who are going to every site […]. Then, to train all the users continually, […] be there to troubleshoot the system with the users, […] monitor the use of the system to make sure it's being used properly and then to monitor that the data is being extracted properly, and […] analyzed properly. That's, I think, where we struggle the most.* [IT or data specialist 8]

Even beyond the specific issue of finding and retaining human resources for point-of-care user support, participants also flagged the overarching issue of insufficient allocation of financial resources to information management overall. They reported that the lack of dedicated funding translates into an inability to upgrade IT infrastructure equipment, acquire or retain skilled human resources, and other necessary investments that are critical for effective information management. One respondent noted that this underresourcing may stem from the iterative process by which electronic information systems often evolve:

> *...But now we need to transition to an electronic system. So, it's already an interactive activity and is being implemented gradually, not completely or integrally. Therefore, there is no specific fund dedicated to it, so that you can say in one or two years, you have 2 or 3 million dollars for it. Because the information system is not like that. It is progressive and with the available resources, you will advance with the strategic direction you want.* [IT or data specialist 9]

The final major operational complexity mentioned by key informants was competing priorities from varied stakeholders, including funders, care providers, and public partners. The key informants identified these competing priorities as significant factors contributing to siloed EMRs, system duplication, and the evolution of an increasingly incoherent and complex information system landscape over time. The competing demands often result in the addition of new parameters to existing systems or the development of parallel systems to meet specific requirements, which can create complexity and hinder the interoperability of information systems. As one respondent noted:

> *Sometimes a partner may ask for a certain parameter, while another partner may ask for a different parameter, and it can be difficult to satisfy both partners.* [IT or data specialist 8]

Respondents also reflected that the growing complexity in the information system landscape is a natural consequence of the

XSL•FO
RenderX

incremental nature of system improvement and the need to continually evolve to meet changing needs:

> *We have some challenges that we face every day, we have an information system that does not necessarily respond to all our needs, so we gradually have to patch it to conform to new demands.* [IT or data specialist 6]

Theme 4: Building Resilient Information Systems

The final theme is focused on participants' insights and aspirations for improvements that could build more robust information systems in the future, overcome current challenges, and deliver greater value to their users.

Prioritizing Continuous System Access

Participants expressed a desire for the development of electronic systems that can be continuously accessed, even when an internet connection is unavailable. Specifically, they hope for systems that can function offline and be connected to a centralized master patient index. This would enable continued use of a system even when there is no internet connection, with local data subsequently synchronized to a broader, cloud-based system once an internet connection becomes available again. The participants emphasized that mitigating connectivity challenges while also enabling integration is key to greater system usability. As one provider noted:

> *One of the requirements I would make is to ask how I can have an offline version of this application. This means that people would be working on the cloud, but there would be synchronization that could be done offline, which would allow for problems with the internet ... or people could work offline but when there is power or internet, it would replicate online.* [Provider 4]

Right-Sizing the Paper Footprint

Key informants hope for the development of an information system that is less dependent on paper. However, their ambitions are diverse in scope. Some contend that a paperless system can be realized if adequate resources are invested, while others suggest that physical files should still function as backup in the event of electronic system failure. The following quotations illustrate these differing perspectives:

> *I think we should keep in mind that we have very few barriers to paperless. [...] I would love to see us come up with a really great interface on tablets, so that we could replace paper. And then, there's no risk of delays and data entry errors and all that. I mean, of course, you can still have data entry errors on tablets. But yeah, I think we could do that.* [IT or data specialist 7]

> *[...] There are always other problems that come up, there are always outages, so we can't guarantee it 100%. The risk that exists is if we can't access the electronic file, then what? Will we never see that file again, or will we just take an old file and put it on paper? That's the dilemma.* [IT or data specialist 8]

Integrated Electronic Medical Records

The respondents commonly expressed the need for a unified patient record system that facilitates the access and sharing of extensive patient information by health care providers in real-time. This objective is believed to enhance patient care coordination and ultimately optimize patient outcomes. Both clinical providers and IT or data specialists shared these aspirations for how better systems might help achieve better service provision to patients. For example:

> *[…] it's important that we have a single electronic system for any patient, because remember the patient's medical history is their story, so if I'm giving the patient care, I have to have access to their history, even with their child's medical history, they were diagnosed with such a problem at such a year, but what was done for them. If they had an allergy to such a medication, I'll know the medication I'm not allowed to give them, they had an allergy at such a time ...* [Provider 4]

Prioritizing Enterprise Resource Planning

Even beyond the common aspirations expressed by most key informants, one participant articulated a more comprehensive future vision, focusing on the necessity of extending system integration beyond patient records. They emphasized the importance of prioritizing the integration of other administrative functions, such as human resources, procurement, finance, and operations. The aim of such integration is to enhance the efficiency of the services that support care delivery within hospitals. This vision and perspective are clearly articulated in the following quotation:

> *Today, if you go to the Purchasing or HR department, there is no application for managing their business processes. There is no application that manages inventory resources, no application that manages purchasing, logistics, and many others. The majority of non-clinical professions do not have applications that ensure the management of their processes. [...] One of the well-understood projects is to come up with an ERP [Enterprise Resource Planning system], a type of software that is fully integrated. [...] And this ERP would have an API [Application Programming Interface] that could allow us to integrate it with other existing applications at the system level.* [IT or data specialist, Haiti 9.]

Discussion

Overarching Findings

The research findings suggest that the long-term sustainability of digital information systems, such as EMRs, in fragile, resource-constrained settings depends not only on modern innovative solutions but on the broader ecosystemic constraints [13,14] and internal complexities in which these systems are embedded. This finding is congruent with a system-theoretical perspective, as it points to the necessity of a holistic approach to analyzing interconnected systems [15-17], in particular the sociotechnical analytical framework, which theorizes the

XSL•FO
RenderX

complex interplay of technology, users, workflows, and institutional environments [15] to understand technological change. It also reveals that, despite many existing challenges, stakeholders perceive strong promise in digital systems to help overcome the serious limitations inherent in paper-based systems and to improve patient care. More importantly, the results of this study suggest several protective strategies that, if implemented well, can increase long-term EMR resilience and enhance their maintainability in such fragile contexts.

Mitigating Multi-Cascading Failures Across Interdependent Systems

The concept of resilience in HISs presupposes, by definition, the ability of the information system to withstand disruptions and recover core functionality after failure [15,18,19]. Although digital system integration offers important potential benefits for the continuity of care, this continuity may be undermined by multi-cascading failures from interdependent systems [20-22]. In resource-constrained settings, inadequate national infrastructure is a recurrent challenge [3,8,13,23-28]. As observed in this study, the literature on electronic health system implementation in comparable contexts consistently highlights the constraints imposed by unreliable interdependent infrastructure [14,23]. In such settings, the occurrence of disturbances affecting interdependent systems, particularly electricity and internet connectivity, is often inevitable. Concurrently, the joint optimization of technical and social parameters is rarely attainable, given the significant resource investments required and the overarching scope of such interventions. From a theoretical perspective, the frequent occurrence of disruptions destabilizes the established resilience theoretical framework distinction of pre- and postdisaster phases and suggests the need for a new perspective that accounts for the state of permanent crisis in such contexts. Thus, there is a need for devising mitigation strategies to strengthen HIS resiliency.

The content analysis of the interview transcripts suggests an implicit consensus among participants on the need to prioritize enhancing the robustness and redundancy attributes of HIS resilience. Participants indicate that resiliency can be strengthened by building redundancy into power sources on premises to mitigate reliance on unstable public uses and diversifying connectivity sources including the adoption of emerging technologies, such as Starlink low-earth orbit satellite internet, which may constitute a viable solution in infrastructure-constrained settings. Respondents also described the need to strengthen the robustness attribute through proper routine maintenance of on-premises IT and power infrastructure, as well as through a hybrid information system topology with right-sized paper footprint serving as a buffer component to EMR failure. The literature on interdependent infrastructure resilience consistently identifies redundancy and robustness as foundational attributes for improving system resilience [29-31].

Although redundancy in power sources remains essential, viable alternatives to many on-premises IT infrastructure do exist, particularly through the adoption of cloud-based solutions. However, this transition comes with an important trade-off: the reduction in costs associated with IT infrastructure maintenance and specialized IT staff salaries comes with an increased dependency on reliable electricity and internet connectivity.

Perceived Promise of Electronic Systems

A key component of the perceived promise of electronic systems stems from the obvious untenability of paper-based information systems [32,33]. The limitations of physical storage, as paper files necessitate space that cannot be perpetually expanded [34,35], contribute to suboptimal management of patient records, with potentially consequential implications for the continuity and quality of care delivered to patients. Participants suggested that the gradual dematerialization of paper records through systematic scanning may offer a pragmatic pathway toward reducing the paper footprint of the health information system. Extending this perspective, contemporary AI-enabled cloud solutions, such as Azure AI Document Intelligence and native AI document parser in the Databricks platform, may provide relatively affordable and scalable means for accelerating the digitization, extraction, and processing of information contained in scanned records [36,37]. However, the practicality of such solutions in resource-constrained settings hinges on the ability to fully transition to modern, cloud-based enterprise resource planning infrastructure, a shift that may increase dependence on unstable interdependent systems: electricity and connectivity. A paradox emerges, wherein paper-based systems are perceived simultaneously as a problem to be addressed through the transition to EMR and as a necessary stop-gap solution during gaps in power or connectivity.

Moreover, another perceived promise of modern EMRs lies in their ability to support continuity of care along the patients' care pathway across time and space. Central to this process is EMR "interoperability," defined as the capacity of diverse EMR subsystems within the broader HIS macro-system to exchange, interpret, and use information coherently and seamlessly. Interoperability therefore constitutes a foundational mechanism of integration, through which otherwise distinct digital components are functionally connected to support coordinated system performance and broader health system objectives. In the context of the 2 public health facilities included in this study, the semantic layer of EMRs aligns, to some extent, with known standards such as SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms) for clinical terminology and LOINC (Logical Observation Identifiers Names and Codes) for laboratory testing, but the inability to uniquely identify patients remains a significant barrier to interoperability. Therefore, achieving effective interoperability depends on an underlying MPI that can reliably establish and maintain patients' identities across time and place, thereby supporting EMR integration across vertical programs and among health care institutions over time [13,32,38-40].

Existing literature originating from studies in low-resource settings emphasizes both the significance of an MPI in achieving interoperability and the challenges in making MPI a reality. For instance, a study conducted in Nigeria highlights the challenges associated with the development of a reliable MPI, as no current strategy satisfactorily fulfills the requisite attributes (uniqueness, unchanging, uncontroversial, inexpensive, ubiquitous, and uncomplicated) for dependability [13,32,38-40]. These concerns

XSL•FO
RenderX

are particularly pertinent in the context of Haiti, where the national identification system is unreliable and incomplete, especially among the most impoverished citizens—a common challenge in resource-constrained contexts [26,38,41].

While it is easy to imagine "digital transformation" as a linear progression from paper-based systems to basic electronic systems and ultimately to electronic systems with increasingly sophisticated functionality, this research found that "legacy" systems create substantial internal complexities that can in fact be an obstacle to improving HISs. These legacy systems or vertical, disease-specific EMRs tend to dominate HISs in LMICs [42]. The complexities engendered by these legacy systems stem from concerns related to their initial design and underlying technologies, which cannot be adequately remedied through patchwork modifications. Simultaneously, a comprehensive revamping of these legacy systems necessitates substantial financial and human resources [6,43] that may not be immediately available and could be perceived as a risky place to invest.

Protective Strategies

The results of this study suggest several protective strategies that could help implementers increase EMR resilience. First, a strong focus on the integration of systems is a strategy that can address the internal complexities related to legacy systems [38,44]. While full-scale replacement of siloed systems—which are often managed by disparate partner organizations—may not be feasible, investment in initiatives to connect systems through application programming interfaces, streamline terminologies, or otherwise facilitate data exchange and interpretability between systems may help users access the information they need and reduce the need for error-prone duplicate data collection across multiple platforms [8].

A second protective factor involves a stronger focus on requirements for EMRs to ensure better alignment with existing clinical workflows, support clinicians [45,46], and limit the duplicated data collection burden that poorly align with patient care processes [47,48]. This necessitates prioritizing viable solutions that address existing challenges without introducing further complications. As aptly articulated by one participant: "when you implement a system and tell the 'provider' that I am simplifying your life, you cannot invent a solution that complicates life [...]."

Finally, the findings underscore the importance of adopting emerging technological innovations that may alter the conditions for EMR implementation in sudden and positive ways. Innovations such as low-Earth-orbit satellite internet service may help mitigate persistent infrastructure constraints characteristic of LMICs.

This study bridges 2 strands of sociotechnical literature: one concerned with the historical, contingent, and path-dependent dynamics of technological transitions and another focused on the configurations of conditions that enable such transitions toward more sociotechnical equilibria [49]. Drawing on empirical evidence of EMR implementation in Haiti, the findings show that digital HIS transformation in fragile and resource-constrained settings cannot be understood as a discrete act of technology adoption. Rather, it is a sociotechnical transition shaped by ecosystemic constraints, internal complexities, and resilience strategies that determine whether EMRs can become durable components of care delivery.

Limitations

This study focused on EMR implementation at 2 public health care facilities supported by ZL in Haiti. Although the findings provide important insights into EMR sustainability in resource-constrained settings, they should be applied with caution when considering applicability and generalizability to other health care institutions in Haiti or to other LMIC contexts. ZL-supported facilities may benefit from additional technical, financial, infrastructural, and organizational resources that are not consistently available in public facilities funded primarily through local or national governmental mechanisms.

Furthermore, the research did not examine the perspectives of critical stakeholders from the Haitian Ministry of Health, who could provide critical insights into how donor-state dynamics shape broader health information policies, governance structures, and priorities related to HIS. This is a critical gap that affects the generalizability of the findings.

Finally, community outreach is increasingly becoming a cornerstone of care delivery in LMICs where access to facility-based care remains limited [50-53]. This study did not examine existing information systems used by community health programs, nor did it assess their potential role in improving the continuity of care through the linking of facility-based care with community-level service delivery.

Conclusions

This study adds to a growing base of evidence shedding light on key factors that support or endanger the success of digital transformation for HISs in resource-constrained settings. The study findings underscore that the successful implementation of EMR for improved clinical care necessitates a comprehensive strategy, considering both the ecosystemic constraints at the national level and the internal complexities within the existing system implementation contexts. It also highlights a set of protective strategies that sustain and improve EMR resilience in fragile and resource-constrained settings. These strategies include prioritizing integration and interoperability between systems, building redundancy to mitigate cascading failures derived from other interdependent systems, strengthening staffing to support system use, right-sizing the paper footprint (eg, improving handwriting-to-text scanning and digitization solutions), and embracing technological innovations.

XSL•FO
RenderX

Acknowledgments

We are thankful to the faculty of the Global Health Delivery Master's Program at Harvard Medical School for their exceptional teaching and inspiration. In particular, we offer our heartfelt thanks to Professor Hannah Gilbert, Professor Mary-Kay Smith Fawzi, Professor Bethany Hedt-Gauthier, Professor Joia Mukherjee, Christina Thompson-Lively, and Nadeem Kasmani for their unwavering support and invaluable guidance throughout this journey. We are especially grateful to Elizabeth Campa, whose enduring support was instrumental in bringing this enterprise to fruition.

To the unforgettable Dr. Paul Farmer—your legacy lives on.

During the preparation of this manuscript, the authors used ChatGPT developed by OpenAI to assist with phrasing suggestions and grammar in the introduction and discussion sections. The authors take full responsibility for the content, originality, accuracy, interpretations, and the integrity of the findings presented in the manuscript.

Funding

We extend our deepest gratitude to the Department of Global Health and Social Medicine at Harvard University for generously funding this research project. This work was conducted with support from the Master of Medical Sciences in GlobalHealth Delivery program of Harvard Medical School Department of Global Health and Social Medicine and financial contributions from Harvard University and the Ronda Stryker and William Johnston MMSc Fellowship in Global Health Delivery. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University and its affiliated academic health care centers.

Authors' Contributions

BDI, the principal investigator of this publication, conducted the work under the supervision of AM. All coauthors reviewed the content, provided feedback, and approved the final manuscript prior to publication.

Conflicts of Interest

None declared.

References

1. Design and implementation of health information systems. : World Health Organization; 2000 URL: https://apps.who.int/iris/handle/10665/42289 [accessed 2021-10-24]
2. Chu A, Phommavong C, Lewis J, Braa J, Senyoni W. Applying ICT to health information systems (HIS) in low resource settings: implementing DHIS2 as an integrated health information platform in Lao PDR. In: Information and Communication Technologies for Development: Springer; 2017. [doi: 10.1007/978-3-319-59111-7_44]
3. Ngugi PN, Were MC, Babic A. Users' perception on factors contributing to electronic medical records systems use: a focus group discussion study in healthcare facilities setting in Kenya. BMC Med Inform Decis Mak 2021 Dec 26;21(1):362. [doi: 10.1186/s12911-021-01737-x] [Medline: 34955098]
4. Al-Rawajfah O, Tubaishat A. Barriers and facilitators to using electronic healthcare records in Jordanian hospitals from the nurses' perspective: a national survey. Inform Health Soc Care 2019 Jan;44(1):1-11. [doi: 10.1080/17538157.2017.1353998] [Medline: 28829654]
5. Huang C, Koppel R, McGreevey JD, Craven CK, Schreiber R. Transitions from one electronic health record to another: challenges, pitfalls, and recommendations. Appl Clin Inform 2020 Oct;11(5):742-754. [doi: 10.1055/s-0040-1718535] [Medline: 33176389]
6. Gyamfi A, Mensah KA, Oduro G, Donkor P, Mock CN. Barriers and facilitators to electronic medical records usage in the emergency centre at Komfo Anokye Teaching Hospital, Kumasi-Ghana. Afr J Emerg Med 2017 Dec;7(4):177-182. [doi: 10.1016/j.afjem.2017.05.002] [Medline: 30456135]
7. deRiel E, Puttkammer N, Hyppolite N, et al. Success factors for implementing and sustaining a mature electronic medical record in a low-resource setting: a case study of iSanté in Haiti. Health Policy Plan 2018 Mar 1;33(2):237-246. [doi: 10.1093/heapol/czx171] [Medline: 29253138]
8. Developing a strategic program for Chilean health information technology: environmental scan and key informant interviews. : Rand Corporation; 2016 URL: http://www.rand.org/pubs/research_reports/RR1358z1.html [accessed 2026-06-25]
9. Braa J, Hanseth O, Heywood A, Mohammed W, Shaw V. Developing health information systems in developing countries: the flexible standards strategy. MIS Q 2007 Jun 1;31(2):381-402. [doi: 10.2307/25148796]
10. Luna D, Almerares A, Mayan JC 3rd, González Bernaldo de Quirós F, Otero C. Health informatics in developing countries: going beyond pilot practices to sustainable implementations: a review of the current challenges. Healthc Inform Res 2014 Jan;20(1):3-10. [doi: 10.4258/hir.2014.20.1.3] [Medline: 24627813]
11. Jazayeri D, Farmer P, Nevil P, Mukherjee JS, Leandre F, Fraser HSF. An Electronic Medical Record system to support HIV treatment in rural Haiti. AMIA Annu Symp Proc 2003;2003:878. [Medline: 14728383]

XSL•FO
RenderX

12. Wolfe BA, Mamlin BW, Biondich PG, et al. The OpenMRS system: collaborating toward an open source EMR for developing countries. AMIA Annu Symp Proc 2006;2006:1146. [Medline: 17238765]
13. Were MC, Savai S, Mokaya B, et al. mUzima Mobile Electronic Health Record (EHR) system: development and implementation at scale. J Med Internet Res 2021 Dec 14;23(12):e26381. [doi: 10.2196/26381] [Medline: 34904952]
14. Akhlaq A, McKinstry B, Muhammad KB, Sheikh A. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review. Health Policy Plan 2016 Nov;31(9):1310-1325. [doi: 10.1093/heapol/czw056] [Medline: 27185528]
15. Tong W, Li HX, Nasirzadeh F, Yao F, Ji Y. Resilience of interdependent infrastructure networks: review and future directions. Int J Crit Infrastruct Prot 2025 Dec;51:100793. [doi: 10.1016/j.ijcip.2025.100793]
16. What is systems theory? : Çağ University; 2017 URL: https://www.cag.edu.tr/uploads/site/lecturer-files/what-is-systems-theory-Aic1.pdf [accessed 2026-06-25]
17. Arkadiusz M. Global theories of archaeology. In: Encyclopedia of Archaeology, 2nd edition: Academic Press; 2024:10-19. [doi: 10.1016/B978-0-323-90799-6.00199-3]
18. Ouyang M. A mathematical framework to optimize resilience of interdependent critical infrastructure systems under spatially localized attacks. Eur J Oper Res 2017 Nov;262(3):1072-1084. [doi: 10.1016/j.ejor.2017.04.022]
19. Fang YP, Zio E. An adaptive robust framework for the optimization of the resilience of interdependent infrastructures under natural hazards. Eur J Oper Res 2019 Aug;276(3):1119-1136. [doi: 10.1016/j.ejor.2019.01.052]
20. Valdez LD, Shekhtman L, La Rocca CE, et al. Cascading failures in complex networks. J Complex Netw 2020 Apr 1;8(2):cnaa013. [doi: 10.1093/comnet/cnaa013]
21. Sharma N, Nocera F, Gardoni P. Classification and mathematical modeling of infrastructure interdependencies. Sustain Resil Infrastruct 2021 Mar 4;6(1-2):4-25. [doi: 10.1080/23789689.2020.1753401]
22. Wells EM, Boden M, Tseytlin I, Linkov I. Modeling critical infrastructure resilience under compounding threats: a systematic literature review. Prog Disaster Sci 2022 Oct;15:100244. [doi: 10.1016/j.pdisas.2022.100244]
23. Mugisha M, Umubyeyi A, Ngenzi JL, et al. Exploring user experience and use of the OpenMRS HIV electronic medical record in primary healthcare facilities of Rwanda: a qualitative study. JMIR Preprints. Preprint posted online on Oct 24, 2023. [doi: 10.2196/preprints.53895]
24. Mkalira Msiska KE, Kumitawa A, Kumwenda B. Factors affecting the utilisation of electronic medical records system in Malawian central hospitals. Malawi Med J 2017 Sep;29(3):247-253. [doi: 10.4314/mmj.v29i3.4] [Medline: 29872515]
25. Jawhari B, Keenan L, Zakus D, et al. Barriers and facilitators to Electronic Medical Record (EMR) use in an urban slum. Int J Med Inform 2016 Oct;94:246-254. [doi: 10.1016/j.ijmedinf.2016.07.015] [Medline: 27573333]
26. Woldemariam MT, Jimma W. Adoption of electronic health record systems to enhance the quality of healthcare in low-income countries: a systematic review. BMJ Health Care Inform 2023 Jun;30(1):e100704. [doi: 10.1136/bmjhci-2022-100704] [Medline: 37308185]
27. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
28. Jawhari B, Ludwick D, Keenan L, Zakus D, Hayward R. Benefits and challenges of EMR implementations in low resource settings: a state-of-the-art review. BMC Med Inform Decis Mak 2016 Sep 6;16(1):116. [doi: 10.1186/s12911-016-0354-8] [Medline: 27600269]
29. Bruneau M, Chang SE, Eguchi RT, et al. A framework to quantitatively assess and enhance the seismic resilience of communities. Earthq Spectra 2003 Nov;19(4):733-752. [doi: 10.1193/1.1623497]
30. Baroud H, Barker K, Ramirez-Marquez JE, Rocco CM. Inherent costs and interdependent impacts of infrastructure network resilience. Risk Anal 2015 Apr;35(4):642-662. [doi: 10.1111/risa.12223] [Medline: 24924523]
31. Xu W, Hong L, He L, Wang S, Chen X. Supply-driven dynamic inoperability input-output price model for interdependent infrastructure systems. J Infrastruct Syst 2011 Dec;17(4):151-162. [doi: 10.1061/(ASCE)IS.1943-555X.0000058]
32. Shekar M, Otto K. ICTs for HEALTH in Africa. : World Bank; 2014 URL: https://openknowledge.worldbank.org/server/api/core/bitstreams/7ed146bd-ce46-5d3e-a73c-17390a661379/content [accessed 2026-06-25]
33. Hagel C, Paton C, Mbevi G, English M, Clinical Information Network (CIN) information systems interest group. Data for tracking SDGs: challenges in capturing neonatal data from hospitals in Kenya. BMJ Glob Health 2020;5(3):e002108. [doi: 10.1136/bmjgh-2019-002108] [Medline: 32337080]
34. Leza D, Phiri J. Challenges of medical records interoperability in developing countries: a case study of the University Teaching Hospital in Zambia. Int J Adv Comput Sci Appl 2019;10(1). [doi: 10.14569/IJACSA.2019.0100171]
35. Abdulkadir A, Aiyedun T, Shoretire K, Danladi A, Anka M, Ologunde K. Paper-based medical records: the challenges and lessons learned from studying obstetrics and gynaecological post-operation records in a Nigerian hospital. TAF Prev Med Bull 2010 Oct;9(5):427-432.
36. Zhang C. Adaptive retrieval-augmented question answering for efficient multi-complexity queries in intelligent systems. Presented at: 2025 5th International Symposium on Artificial Intelligence and Big Data (AIBDF); Dec 26-28, 2025. [doi: 10.1109/AIBDF67964.2025.11440815]
37. Ke W, Zheng Y, Li Y, et al. Large language models in document intelligence: a comprehensive survey, recent advances, challenges, and future trends. ACM Trans Inf Syst 2026 Jan 31;44(1):1-64. [doi: 10.1145/3768156]

XSL•FO
RenderX

38. Jayathissa W, Dissanayake VHW, Hewapathirana R. Developing a functional prototype master patient index (MPI) for interoperability of e-health systems in Sri Lanka. arXiv. Preprint posted online on Mar 1, 2018. [doi: 10.48550/arXiv.1803.00451]
39. Chukwu E, Ekong I, Garg L. Scaling up a decentralized offline patient ID generation and matching algorithm to accelerate universal health coverage: insights from a literature review and health facility survey in Nigeria. Front Digit Health 2022;4:985337. [doi: 10.3389/fdgth.2022.985337] [Medline: 36158996]
40. Thorell L, Molin JD, Fyfe J, Hone S, Lwin SM. Working towards a master patient index and unique identifiers to improve health systems: the example of Myanmar. WHO South East Asia J Public Health 2019 Sep;8(2):83-86. [doi: 10.4103/2224-3151.264851] [Medline: 31441442]
41. Ariffin NABN, Ismail AB, Kadir IKA, Kamal JIA. Implementation of electronic medical records in developing countries: challenges & barriers. Int J Acad Res Prog Educ Dev 2018 Aug;7(3):187-199. [doi: 10.6007/IJARPED/v7-i3/4358]
42. Kumar M, Mostafa J. Research evidence on strategies enabling integration of electronic health records in the health care systems of low- and middle-income countries: a literature review. Int J Health Plann Manage 2019 Apr;34(2):e1016-e1025. [doi: 10.1002/hpm.2754] [Medline: 30762907]
43. Ciccone EJ, Tilly AE, Chiume M, et al. Lessons learned from the development and implementation of an electronic paediatric emergency and acute care database in Lilongwe, Malawi. BMJ Glob Health 2020 Jul;5(7):e002410. [doi: 10.1136/bmjgh-2020-002410] [Medline: 32675067]
44. Higman S, Dwivedi V, Nsaghurwe A, et al. Designing interoperable health information systems using Enterprise Architecture approach in resource-limited countries: a literature review. Int J Health Plann Manage 2019 Jan;34(1):e85-e99. [doi: 10.1002/hpm.2634] [Medline: 30182517]
45. Pugh CM. Electronic health records, physician workflows and system change: defining a pathway to better healthcare. Ann Transl Med 2019 Mar;7(Suppl 1):S27. [doi: 10.21037/atm.2019.01.83] [Medline: 31032307]
46. Zheng K, Ratwani RM, Adler-Milstein J. Studying workflow and workarounds in electronic health record-supported work to improve health system performance. Ann Intern Med 2020 Jun 2;172(11 Suppl):S116-S122. [doi: 10.7326/M19-0871] [Medline: 32479181]
47. Tun SYY, Madanian S. Clinical information system (CIS) implementation in developing countries: requirements, success factors, and recommendations. J Am Med Inform Assoc 2023 Mar 16;30(4):761-774. [doi: 10.1093/jamia/ocad011] [Medline: 36749093]
48. Kabukye JK, de Keizer N, Cornet R. Elicitation and prioritization of requirements for electronic health records for oncology in low resource settings: a concept mapping study. Int J Med Inform 2020 Mar;135:104055. [doi: 10.1016/j.ijmedinf.2019.104055] [Medline: 31877404]
49. Spijkerboer RC, Turhan E, Roos A, et al. Out of steam? A social science and humanities research agenda for geothermal energy. Energy Res Soc Sci 2022 Oct;92:102801. [doi: 10.1016/j.erss.2022.102801]
50. Lassi ZS, Bhutta ZA. Community-based intervention packages for reducing maternal and neonatal morbidity and mortality and improving neonatal outcomes. Cochrane Database Syst Rev 2015 Mar 23;2015(3):CD007754. [doi: 10.1002/14651858.CD007754.pub3] [Medline: 25803792]
51. Buthelezi UE, van Rensburg AJ, Moshabela M, et al. A realist perspective on optimizing community health workers' roles and functions to deliver integrated people-centred care. PLoS Glob Public Health 2025;5(9):e0004926. [doi: 10.1371/journal.pgph.0004926] [Medline: 40901841]
52. LeBan K, Kok M, Perry HB. Community health workers at the dawn of a new era: 9. CHWs' relationships with the health system and communities. Health Res Policy Syst 2021 Oct 12;19(Suppl 3):116. [doi: 10.1186/s12961-021-00756-4] [Medline: 34641902]
53. Ijumba P, Doherty T, Jackson D, et al. Effect of an integrated community-based package for maternal and newborn care on feeding patterns during the first 12 weeks of life: a cluster-randomized trial in a South African township. Public Health Nutr 2015 Oct;18(14):2660-2668. [doi: 10.1017/S1368980015000099] [Medline: 25660465]

Abbreviations

EMR: electronic medical record
HIS: health information system
HUM: Hôpital Universitaire de Mirebalais
IRB: institutional review board
LMIC: low- and middle-income country
LOINC: Logical Observation Identifiers Names and Codes
MPI: master patient index
MSPP: Haitian Ministry of Health
PI: principal investigator
SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms
ZL: Zanmi Lasante

XSL•FO
RenderX

Edited by E Mensah; submitted 14.Aug.2025; peer-reviewed by D Chrimes, M Galani, R Ruparel; revised version received 12.May.2026; accepted 17.Jun.2026; published 27.Jul.2026.

Please cite as:
Isaac BD, Michaelis A, Casella Jean-Baptiste M, Viaud LG, Lambert W
Investigating Strategies for Sustaining Modern Health Information Systems for Enhanced Care Delivery in Fragile Resource-Constrained Settings Through Electronic Medical Record Implementation in Haiti: Qualitative Study
Online J Public Health Inform 2026;18:e82396
URL: https://ojphi.jmir.org/2026/1/e82396
doi:10.2196/82396

© Benito D Isaac, Annie Michaelis, Meredith Casella Jean-Baptiste, Loune G Viaud, Wesler Lambert. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 27.Jul.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Global Trends in Phenylketonuria Treatment Research, 2000-2025: Bibliometric Analysis

Sitong Yang[1,2]; Kaichao Song[1,2]; Qingbo Chen[1,2]; Jiandong Jiang[1,2]; Lulu Wang[1,2]

[1]State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 1 Tiantan Xili, Dongcheng District, Beijing, Beijing, China

[2]Beijing Key Laboratory of Technology and Application for Anti-Infective New Drugs Research and Development, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, Beijing, China

Corresponding Author:
Lulu Wang
State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 1 Tiantan Xili, Dongcheng District, Beijing, Beijing, China

Abstract

Background: Phenylketonuria (PKU) is the most common inborn error of amino acid metabolism, and if untreated, leads to severe neurocognitive impairment. Over the past 2 decades, treatment strategies have evolved from strict dietary phenylalanine restriction to include pharmacological therapies such as tetrahydrobiopterin and, more recently, enzyme substitution with pegvaliase. Despite these advances, significant heterogeneity exists in global research priorities, collaboration patterns, and the translation of emerging therapies into clinical practice. A systematic overview of the field's development, thematic shifts, and remaining knowledge gaps is currently lacking.

Objective: This study aimed to provide a systematic bibliometric analysis of global treatment research on PKU from 2000 to 2025. The aim was to quantify publication trends, collaboration patterns, thematic evolution, and research gaps, thereby informing future scientific and clinical directions.

Methods: A search of the Web of Science Core Collection was performed on September 13, 2025. The search initially identified 1877 records. After screening, 1462 English-language articles and reviews were included. Publication trends were analyzed using Microsoft Excel (version 16.101), while VOSviewer 1.6.20 and CiteSpace 6.4R1 were used to visualize country- and institutional-level collaborations, journal networks, keyword co-occurrence, citation bursts, and thematic clusters. Statistical charts were generated with GraphPad Prism 10.2.1.

Results: Annual publication output demonstrated an overall upward trend, peaking in 2022 with 111 publications. The United States led in both publication volume (375 studies) and total citations (11,142 citations), maintaining strong collaborative ties with several European countries, particularly the Netherlands and the United Kingdom. China ranked seventh globally in publication volume, although its citation impact remains comparatively limited. Key institutions, including the University of Groningen and Birmingham Children's Hospital, as well as prominent scholars such as Francjan J van Spronsen and Anita MacDonald, have occupied central positions in the global PKU treatment research collaboration network over the study period. High-frequency and high-centrality keywords, such as "phenylalanine," "dietary treatment," and "tetrahydrobiopterin," highlighted continued emphasis on metabolic control and targeted therapies. Keyword burst analysis revealed a gradual shift from conventional dietary management toward enzyme replacement therapies, investigations of neurocognitive outcomes, and precision medicine–oriented approaches.

Conclusions: Over the past 25 years, PKU treatment research has progressed from foundational dietary interventions to molecular mechanistic studies and individualized therapeutic strategies. Future research should prioritize longitudinal multiomics investigations, targeted metabolic correction technologies, gene-based therapeutic approaches, and enhanced international collaboration, particularly to strengthen diagnosis and management capacities in low- and middle-income regions. Such efforts will be critical to advancing global standards of PKU care.

(Online J Public Health Inform 2026;18:e90419) doi:10.2196/90419

KEYWORDS

phenylketonuria; PKU; bibliometric analysis; metabolic disorder; dietary management; precision medicine

XSL•FO
RenderX

Introduction

Inherited metabolic disorders comprise a group of rare conditions caused by pathogenic gene mutations that result in enzymatic defects and subsequent biochemical disturbances. These disorders can profoundly affect growth and overall quality of life in pediatric patients. Among them, amino acid metabolism defects hold particular importance, with phenylketonuria (PKU) being one of the most prevalent and representative conditions. Since its initial description, PKU has remained a focal topic in medical genetics, pediatrics, and nutritional science, with a global incidence estimated at approximately 1 in 10,000 live births [1].

PKU is an autosomal recessive disorder primarily caused by mutations in the phenylalanine hydroxylase (PAH) gene, which lead to markedly reduced or absent PAH enzyme activity in hepatocytes. This metabolic defect disrupts the normal conversion of phenylalanine to tyrosine, resulting in the pathological accumulation of phenylalanine in the blood, cerebrospinal fluid, and peripheral tissues. Alternative metabolites, such as phenylpyruvic acid, are excreted in urine, giving rise to the condition's name. Elevated phenylalanine concentrations exhibit neurotoxic effects that severely impair brain development and function. Without timely treatment, affected individuals may develop progressive and irreversible intellectual disability, psychiatric and behavioral abnormalities, epilepsy, microcephaly, and other serious neurological impairments [2].

The global implementation of newborn screening programs and improved genetic disease management has substantially altered the natural history of PKU. Early diagnosis combined with lifelong dietary therapy can effectively prevent severe neurodevelopmental impairment [3]. However, as increasing numbers of patients reach adulthood and advanced age, new challenges have emerged, including long-term dietary adherence, psychosocial burdens, maternal PKU management, and adult-onset complications. These issues continue to expand and diversify the scope of PKU research.

Since the beginning of the 21st century, rapid advances in genomics, proteomics, and pharmaceutical development have transformed PKU research into a multidisciplinary knowledge system covering basic biomedical science, clinical pediatrics, nutrition, psychology, and health policy management. Given the dramatic growth of the literature, researchers, clinicians, and policymakers face increasing difficulty in understanding the field's development trajectory, key knowledge structures, and evolving research priorities through traditional narrative reviews alone. Consequently, there is a pressing need for objective, quantitative approaches capable of providing macroscopic insights into the development of PKU research.

Bibliometrics, which applies quantitative statistical methods to academic publications, offers a powerful means of capturing research dynamics, identifying hotspots, mapping collaboration networks, and evaluating scientific influence within a given field [4,5]. Using tools such as co-occurrence analysis, cluster analysis, and collaboration network mapping, bibliometric methods transcend the limitations of individual studies and provide data-driven evidence to support scientific decision-making and strategic planning [6].

The 21st century represents a critical period for optimizing PKU management strategies and achieving breakthroughs in emerging therapies. However, to date, no comprehensive bibliometric analysis has systematically evaluated global PKU research over this period. To address this gap, this study conducts the first bibliometric assessment of PKU-related literature published between 2000 and 2025 in the Web of Science Core Collection (WoSCC). Through this analysis, the study aims to map the scientific knowledge landscape of the field and provide researchers with an objective, systematic reference framework.

Specifically, this study seeks to analyze annual publication trends, geographic distribution, and institutional contributions. Additionally, it aims to identify prolific authors, influential journals, and landmark publications; detect research hotspots and thematic evolution through keyword co-occurrence analysis; and demonstrate scientific collaboration networks to highlight global research linkages. By integrating these elements and discussing future research directions, this work aims to provide a comprehensive, data-driven overview of the development of PKU research. Therefore, this study aimed to systematically evaluate the global landscape of PKU treatment research from 2000 to 2025 using bibliometric methods.

Methods

Data Acquisition and Processing

A comprehensive literature search was performed on September 13, 2025, using the WoSCC, including both SCIE and SSCI editions. WoSCC was selected because it provides standardized bibliographic information and comprehensive citation data and is widely used in bibliometric studies. The search strategy used the following query in the "Topic" field:

(phenylketonuria OR "phenylalanine ketonuria" OR "Folling disease" OR (PKU NOT ("Peking University" OR "Beijing University"))) AND (therap* OR treat* OR manag* OR drug OR pharmacotherap* OR diet* OR sapropterin OR Pegvaliase OR Palynziq OR "case report" OR "clinical study" OR "clinical trial") NOT (screening OR "newborn screen*" OR diagnos* OR detection OR assay OR biomarker).

The publication period was restricted from 2000 to 2025. Because the search was conducted on September 13, 2025, data for 2025 represent only a partial year. This strategy captured the major nomenclature and synonyms associated with PKU while excluding publications related to Peking University. The query focused on literature pertaining to disease management, dietary interventions, pharmacological treatments, and clinical studies, while excluding neonatal screening, diagnostic approaches, and biomarker research.

The initial search yielded 1877 records. Only English-language articles and reviews were retained for analysis. A total of 415 noneligible publications were excluded, including news items (n=4), data papers (n=1), retracted publications (n=3), corrections (n=12), book chapters (n=4), early access articles (n=4), letters (n=23), editorial materials (n=38), meeting

XSL•FO
RenderX

abstracts (n=332), proceedings papers (n=48), book reviews (n=3), retractions (n=2), and reprints (n=1); the document-type categories listed are not mutually exclusive, because some records may be assigned to multiple document types. The final dataset consisted of 1462 publications, including 1272 articles and 190 reviews.

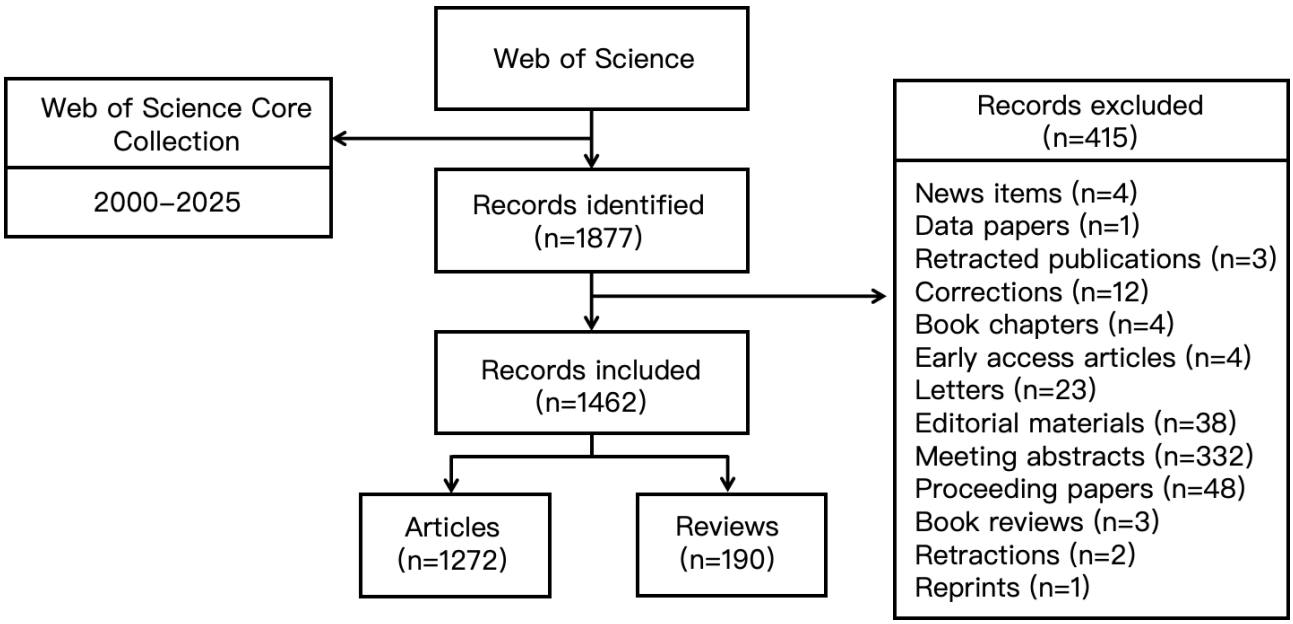
Bibliometric Analysis and Visualization

A bibliometric analysis was conducted using Microsoft Excel (version 16.101), VOSviewer (version 1.6.20; Centre for Science and Technology Studies, Leiden University), and CiteSpace (version 6.4.R1; Drexel University) to generate a comprehensive overview of the research landscape. Microsoft Excel was used to compile annual publication counts from 2000 to 2025, thereby characterizing temporal trends in research productivity. VOSviewer was used to construct visual maps of international collaboration networks, co-authorship relationships, institutional partnerships, and journal co-citation patterns [7]. CiteSpace facilitated journal dual-map overlays, keyword co-occurrence analyses, reference co-citation networks, and burst-detection examinations, with outputs visualized as interpretable network diagrams [8]. The integrated application of these tools provided both quantitative and graphical insights into the intellectual structure, development tendency, and emerging research frontiers within the PKU field. Statistical charts were generated with GraphPad Prism 10.2.1 (GraphPad Software, Inc).

Results

The complete data collection and filtering process is illustrated in Figure 1.

Figure 1. Flowchart for identifying and selecting publications. The document-type categories listed are not mutually exclusive, because some records may be assigned to multiple document types.

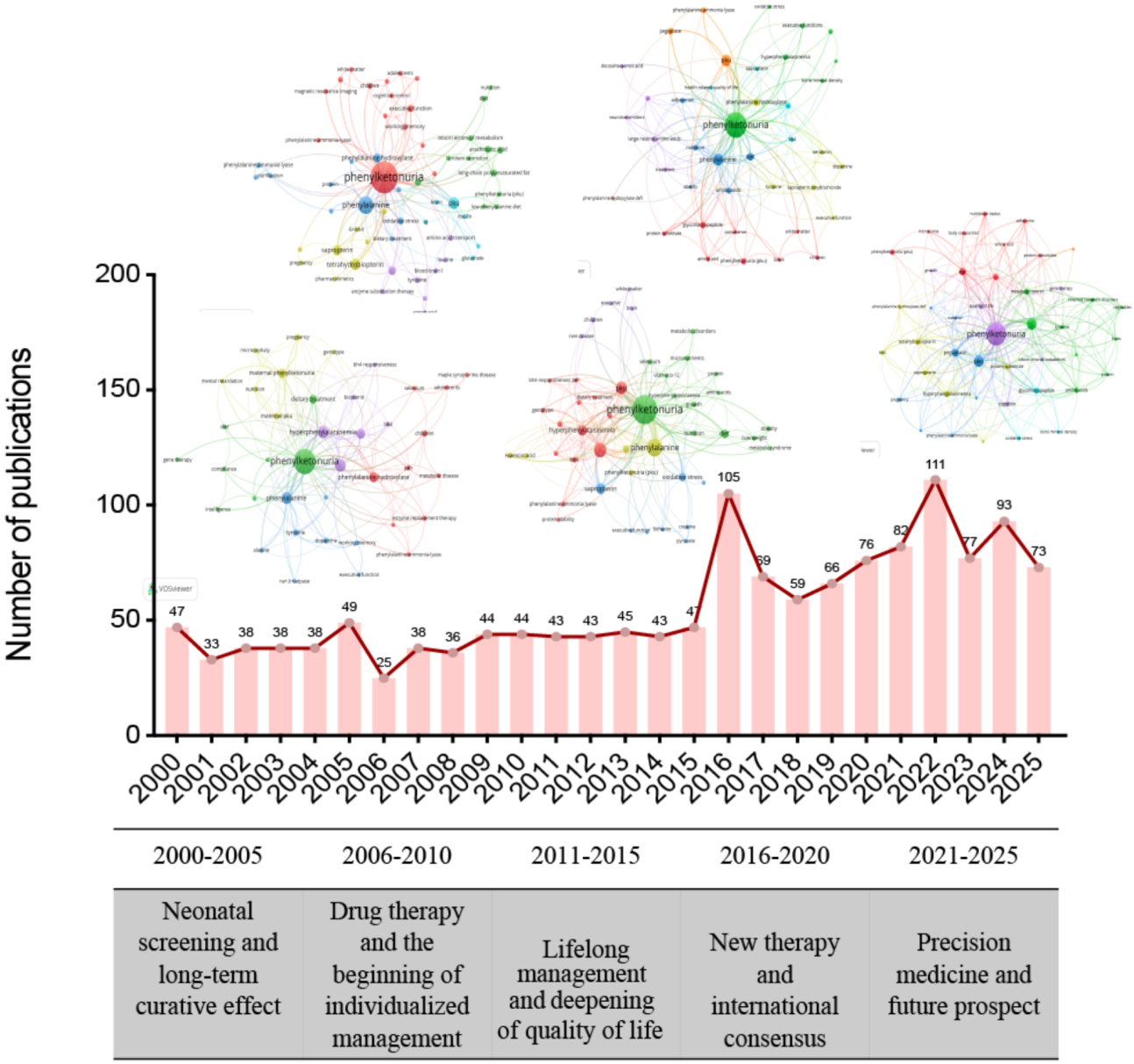


Overview of Research Trends

The temporal distribution of publications provides a clear depiction of research progress in the field of PKU. As illustrated in Figure 2, annual publication output remained relatively stable from 2000 to 2015, with a mean of 40.63 (SD 5.86) publications per year. A modest peak was observed in 2005 (49 publications), followed by a brief decline in 2006 (25 publications). The increase in publication activity coincided with growing interest in sapropterin-based therapies. Although publication numbers fluctuated somewhat in subsequent years, they consistently remained above pre-2015 levels, culminating in a peak of 111 articles in 2022. Despite a slight decline after 2022, annual output demonstrated only minor variation, reflecting sustained scholarly interest and ongoing innovation in this field.

XSL•FO
RenderX

Figure 2. Annual publication volume and 5-year keyword cluster evolution analysis.



In addition to publication trends, Figure 2 presents a comparative keyword analysis using 5-year intervals beginning in 2000, capturing the evolution of research priorities and emerging thematic trends within the field. Keywords were ranked according to both frequency and centrality, with the top 30 keywords for each interval summarized in Table 1. This longitudinal analysis reveals a clear shift in research emphasis: from an early focus on preventing intellectual disability toward improving long-term quality of life, and from managing metabolic abnormalities to developing potentially curative therapeutic strategies, and, more recently, toward emerging approaches such as gene therapy aimed at providing sustained metabolic correction [9,10].

XSL•FO
RenderX

Table . Top 30 keywords across 5-year intervals.

Rank	2000 - 2005		2006-2010		2011-2015		2016-2020		2021-2025	
	Keywords	Occurrences, n	Keywords	Occurrences, n	Keywords	Occurrences, n	Keywords	Occurrences, n	Keywords	Occurrences, n
1	Phenylketonuria	89	Phenylketonuria	78	Phenylketonuria	125	Phenylketonuria	150	Phenylketonuria	234
2	Hyperphenylalaninemia	22	Phenylalanine	24	Phenylalanine	47	Phenylalanine	34	PKU	58
3	PKU	22	PKU	11	Tetrahydrobiopterin	30	PKU	30	Phenylalanine	57
4	Phenylalanine	19	Tetrahydrobiopterin	10	PKU	24	Diet	15	Pegvaliase	30
5	Phenylalanine; hydroxylase	14	Sapropterin	8	Sapropterin	18	Phenylalanine hydroxylase	12	Diet	25
6	Tetrahydrobiopterin	14	Phenylalanine hydroxylase	7	Hyperphenylalaninemia	16	Glycomacropeptide	11	Hyperphenylalaninemia	17
7	Maternal phenylketonuria	12	Hyperphenylalaninemia	6	Diet	11	Pegvaliase	10	Inborn errors of metabolism	17
8	Dietary treatment	11	Arachidonic acid	4	Oxidative stress	10	Amino acids	9	Glycomacropeptide	16
9	Children	6	Docosahexaenoic acid	4	Tyrosine	9	Phenylketonuria (PKU)	9	Metabolic control	16
10	Maternal PKU	6	Executive function	4	Phenylketonuria (PKU)	6	Hyperphenylalaninemia	8	Protein substitute	14
11	Tyrosine	6	Gene therapy	4	BH4	5	Nutrition	8	Phenylketonuria (pku)	13
12	Dopamine	5	Long-chain polyunsaturated fatty acids	4	Docosahexaenoic acid	5	Phenylalanine ammonia lyase	8	Tetrahydrobiopterin	13
13	Pregnancy	5	Oxidative stress	4	Growth	5	Tetrahydrobiopterin	8	Amino acids	12
14	Alanine	4	Amino acid transport	3	Micronutrients	5	Adherence	7	Sapropterin	11
15	Blood-brain barrier	4	Children	3	Phenylalanine hydroxylase	5	Executive functions	7	Cognition	10
16	Enzyme replacement therapy	4	Diet	3	Children	4	Inborn error of metabolism	7	Quality of life	10
17	Microcephaly	4	Diffusion tensor imaging	3	Genotype	4	Large neutral amino acids	7	Adherence	9
18	Nutrition	4	Inborn errors of metabolism	3	Inborn error of metabolism	4	Quality of life	7	Pregnancy	9
19	PAH	4	Phenylalanine ammonia lyase	3	Inborn errors of metabolism	4	Amino acid	6	Tyrosine	9

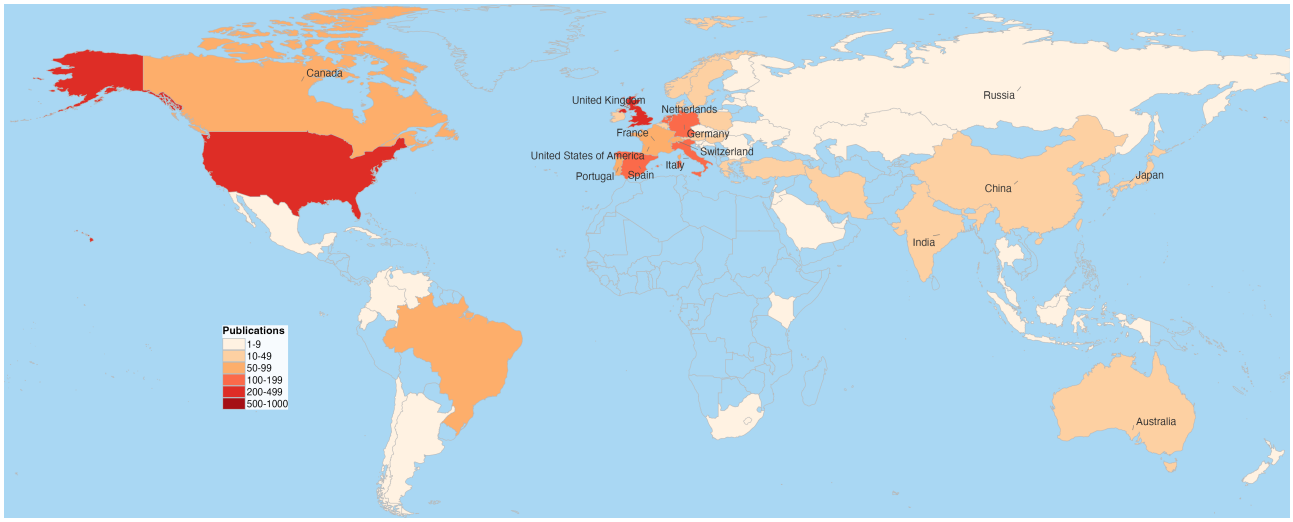
XSL•FO
RenderX

Rank	2000 - 2005		2006-2010		2011-2015		2016-2020		2021-2025	
	Keywords	Occurrences, n	Keywords	Occurrences, n	Keywords	Occurrences, n	Keywords	Occurrences, n	Keywords	Occurrences, n
20	Working memory	4	Phenylketonurias	3	Metabolic disorders	4	Dopamine	6	Inherited metabolic disorders	8
21	Adolescents	3	Tyrosine	3	Nutrition	4	Sapropterin dihydrochloride	6	Nutritional status	8
22	BH4	3	Working memory	3	Obesity	4	Tyrosine	6	Oxidative stress	8
23	BH4 responsiveness	3	6R-BH4	2	Overweight	4	BH4	5	BH4	7
24	Biopterin	3	Activated carbon	2	Phenylalanine ammonia lyase	4	Bone mineral density	5	Body composition	7
25	Compliance	3	Adeno-associated virus	2	Quality of life	4	Children	5	Gene therapy	7
26	Diet	3	Adolescents	2	Sapropterin dihydrochloride	4	Docosahexaenoic acid	5	Phenylalanine hydroxylase deficiency	7
27	Executive function	3	Amino acid	2	Selenium	4	Protein substitute	5	Bone mineral density	6
28	Gene therapy	3	Blood-brain barrier	2	Amino acids	3	Sapropterin	5	Inborn error of metabolism	6
29	Genotype	3	Brain	2	Behavior	3	Serotonin	5	Nutrition	6
30	Inborn errors of metabolism	3	Cell transplantation	2	BH4-responsiveness	3	Brain	4	Phenylalanine ammonia lyase	6

Country Analysis

A total of 64 countries contributed to the global publication output on PKU. As shown in Table 2, the top 10 productive countries were ranked based on their scientific output. The United States dominated the field with 375 publications and 11,142 citations, confirming its central role in both research productivity and scholarly influence. The United Kingdom ranked second in publication volume (185 publications) and citation impact (4269 citations), and also demonstrated the highest level of international collaboration, reflecting its strong integration within the global scientific community. The worldwide distribution of publications is depicted in Figure 3.

Figure 3. Distribution of publications.



XSL•FO
RenderX

Table . Top 10 countries in terms of publication output.

Rank	Country	Documents, n	Citations, n	Total link strength, n
1	United States	375	11,142	188
2	United Kingdom	185	4269	274
3	Germany	144	4004	247
4	Italy	121	2806	182
5	The Netherlands	103	3281	199
6	Spain	85	2140	169
7	Switzerland	72	2560	105
8	China	67	2174	21
9	Canada	58	1931	64
10	Australia	35	1719	16

Given the higher prevalence of PKU among Caucasian populations, European countries exhibited strong representation in the literature. Germany, the Netherlands, Italy, and Switzerland ranked third through sixth, with publication counts of 144, 103, 121, and 72 publications, and 4004, 3281, 2806, and 2560 citations, respectively. Notably, Germany displayed the second-highest collaboration intensity, which exceeded even that of the United States.

Annual publication trends for these 10 leading countries are illustrated in Figure 4. VOSviewer was used to construct a geographical visualization of national contributions and collaborative patterns, applying a minimum threshold of 5 publications per country. This criterion yielded 38 countries included in the network analysis. In the resulting visualization, node size corresponds to publication output, whereas link thickness corresponds to the strength of collaborative relationships. Additionally, Figure 5 depicts the temporal evolution of the top 10 countries' publication shares as a percentage of total global output, displayed in 5-year intervals over the 25-year period. The United States consistently maintained the largest share of publications, although its relative proportion declined slightly over time. In contrast, the United Kingdom and Italy exhibited a moderate upward trend in their publication shares.

XSL•FO
RenderX

Figure 4. Map of the national cooperation network.

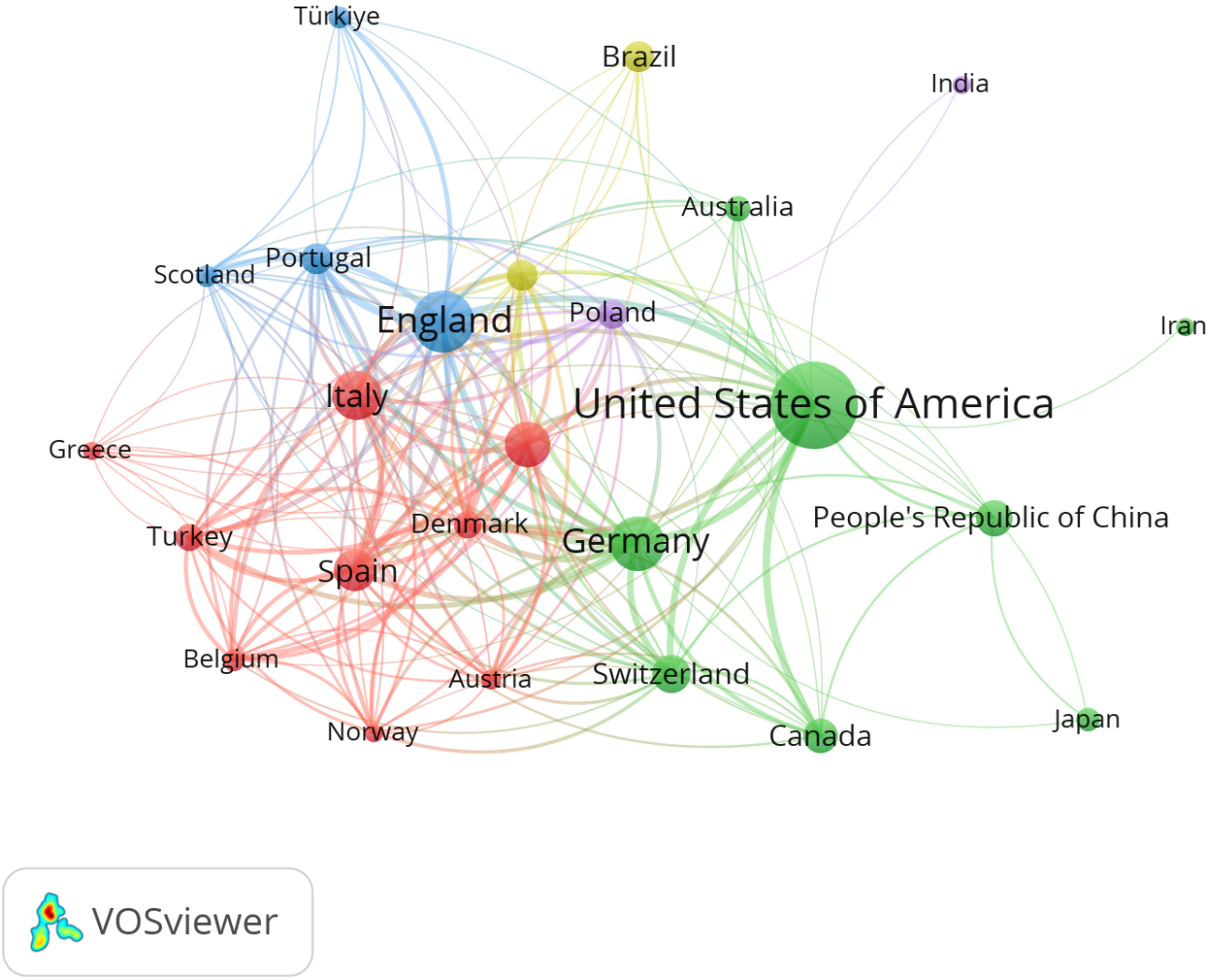
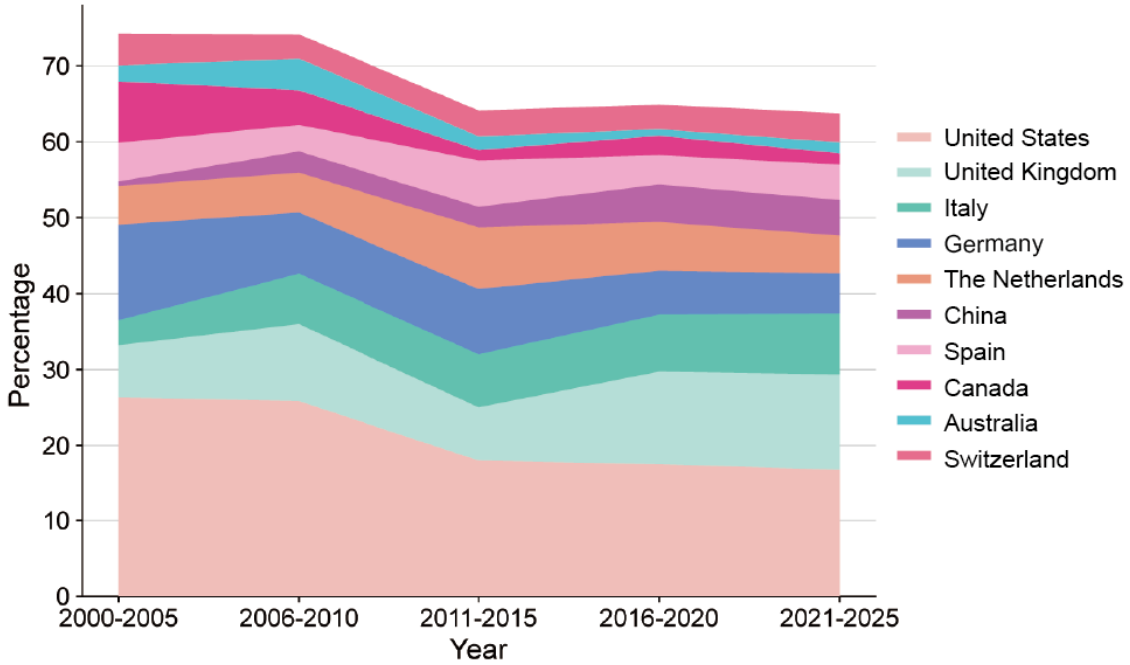


Figure 5. Percentage of publications by country, displayed in 5-year intervals over the 25-year period.



XSL•FO
RenderX

Author Analysis

A total of 5359 researchers contributed to PKU-related publications. Figure 6 presents the author collaboration network, visualizing the collaborative relationships among the most productive and influential researchers in this field. As shown in Table 3, Anita MacDonald from the United Kingdom lead the field with 92 publications and 1613 citations. Their substantial involvement in developing international diagnostic and treatment guidelines, widely regarded as the "gold standard" for PKU management, has resulted in a particularly high citation count [11]. Similarly, Francjan J van Spronsen from the Netherlands, a major contributor to international consensus guidelines, ranks second with 40 publications and 1058 citations, demonstrating extensive collaboration with MacDonald [11,12].

Figure 6. Author collaboration network.

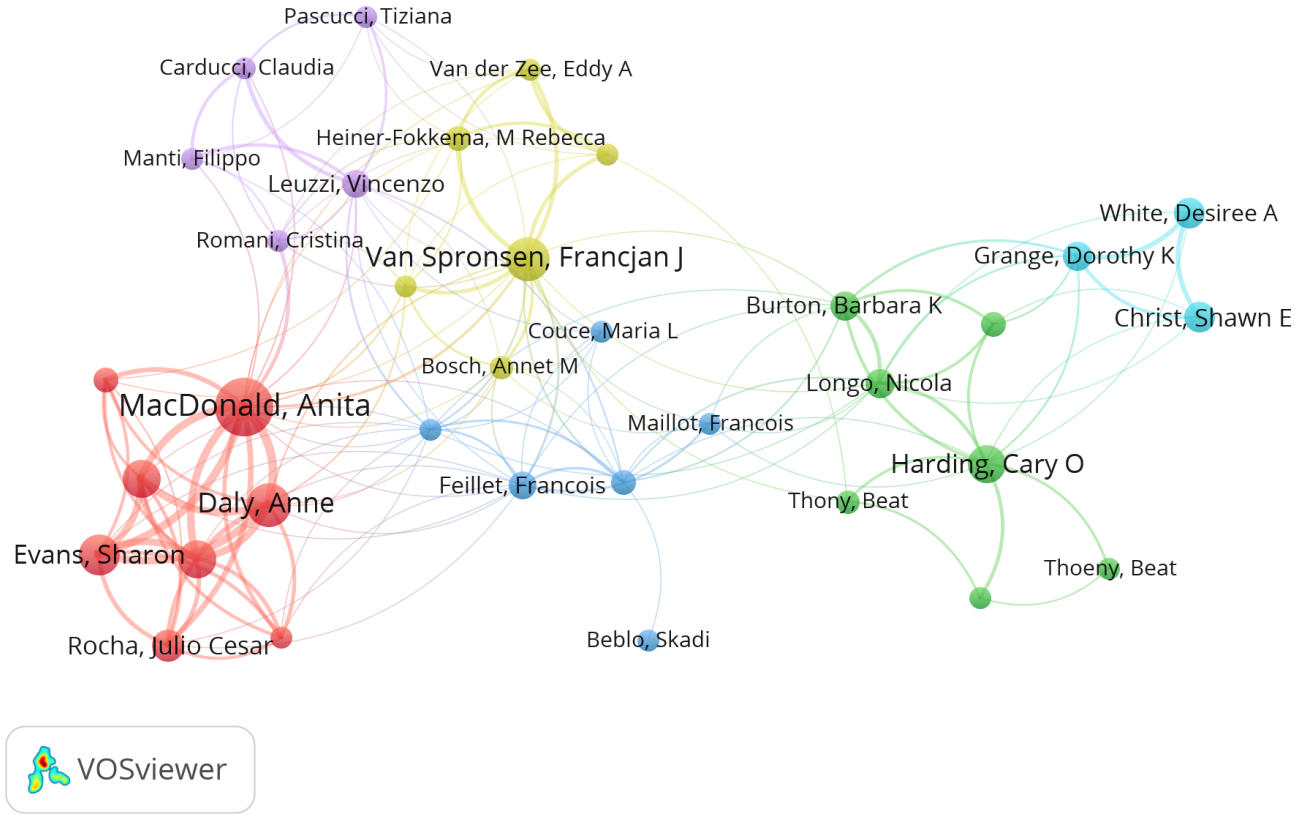


Table . Top 10 authors in terms of publications.

Rank	Author (surname, forename)	Documents, n	Citations, n	Total link strength, n
1	MacDonald, Anita	92	1613	182
2	Van Spronsen, Francjan J	40	1058	64
3	Daly, Anne	36	272	143
4	Harding, Cary O	32	1009	41
5	Evans, Sharon	31	254	132
6	Ashmore, Catherine	28	149	120
7	Pinto, Alex	28	191	123
8	Christ, Shawn E	20	486	24
9	White, Desiree A	20	514	30
10	Burton, Barbara K	19	543	33

Beyond these 2 international leaders, several additional researchers from the United Kingdom, including Anne Daly, Sharon Evans, and Catherine Ashmore, form the core of European research on PKU clinical management. In parallel, US investigators, including Cary O Harding, Shawn E Christ, Desiree A White, and Barbara K Burton, have played key roles in translating foundational scientific discoveries into innovative therapies that have contributed substantially to the development and clinical evaluation of treatments later approved by regulatory agencies, thereby establishing the US central position in PKU research.

XSL•FO
RenderX

Institutional Analysis

A total of 1708 institutions contributed to PKU-related publications. Using a minimum threshold of 13 publications per institution, VOSviewer generated a collaborative network composed of 41 institutional nodes. Figure 7 presents the resulting network map of cooperation among organizations, visualizing the collaborative relationships and clustering patterns of the most productive institutions in this field. Table 4 lists the top 10 institutions ranked by publication volume. Table 4 lists the top 10 institutions ranked by publication volume. The University of Groningen led with 68 publications and 2006 citations and exhibited the highest degree of collaborative connectivity. As PKU is an inherited metabolic disorder frequently managed in pediatric settings, children's hospitals have played an important role in advancing the field. Birmingham Children's Hospital ranked second with 52 publications and 1242 citations.

Figure 7. Network map of cooperation among organizations.

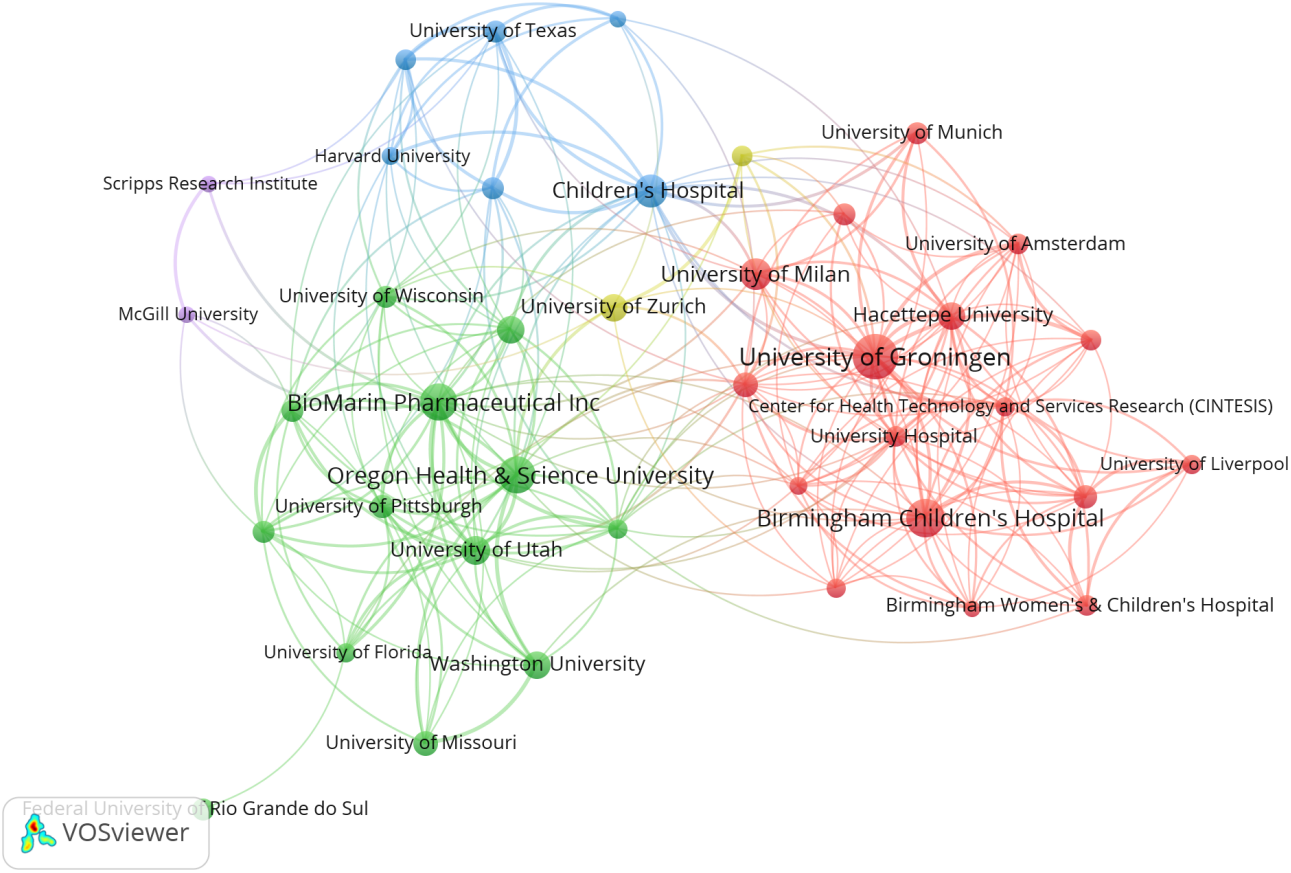


Table . Top 10 institutions by publication output.

Rank	Organization	Documents, n	Citations, n	Total link strength, n
1	University of Groningen	68	2006	102
2	Birmingham Children's Hospital	52	1242	49
3	Oregon Health & Science University	50	1546	89
4	Biomarin Pharmaceutical Inc	48	1548	96
5	Children's Hospital	40	1285	79
6	University of Milan	36	844	31
7	The University of Utah	32	1230	78
8	Hacettepe University	30	558	62
9	University of Zurich	29	1191	24
10	Washington University	29	1036	46

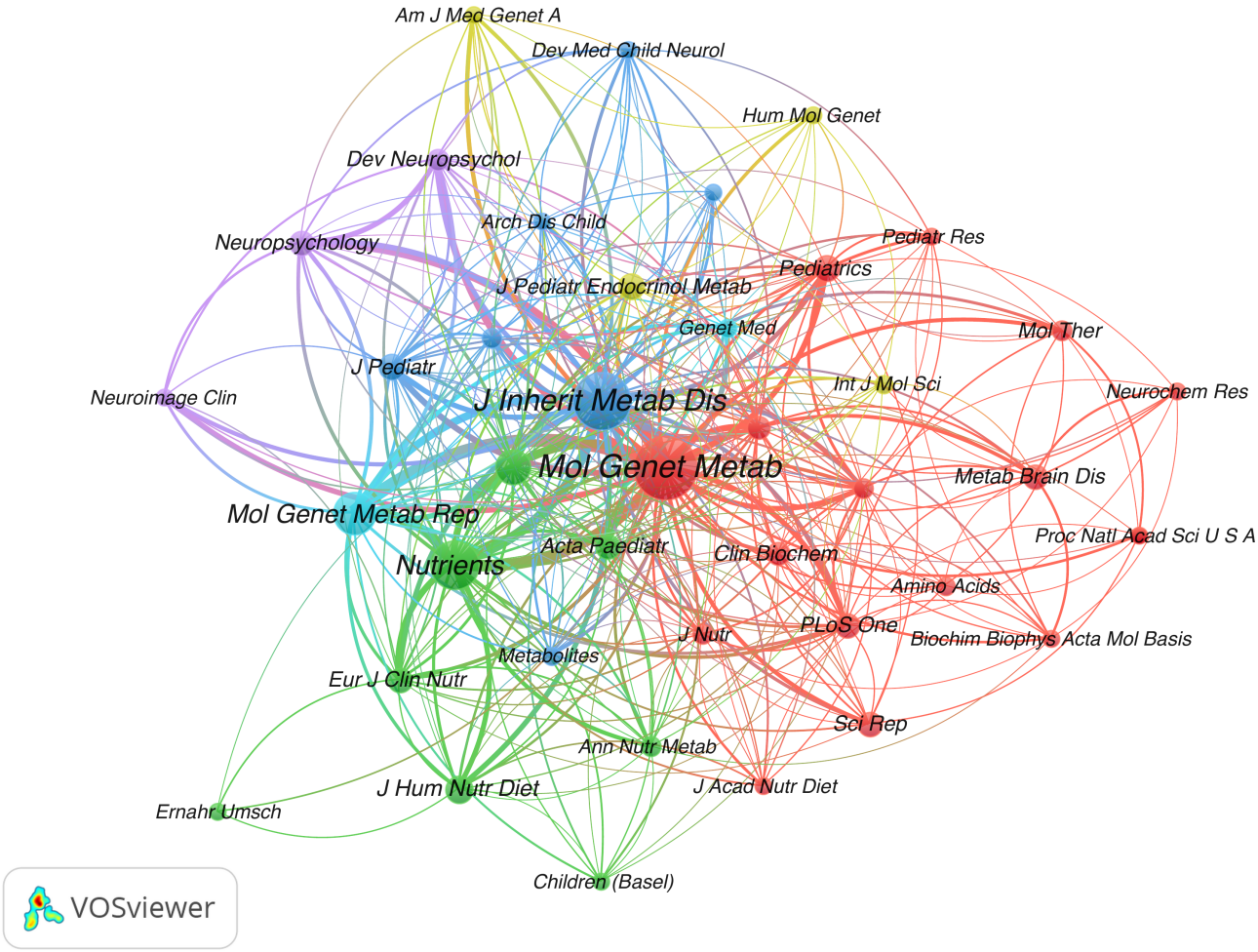
XSL•FO
RenderX

Overall, institutions active in PKU research maintain strong and extensive collaborative relationships, forming a highly cohesive scientific community. The dense interinstitutional linkages observed in the network map indicate a well-integrated research ecosystem that facilitates efficient knowledge exchange and the rapid translation of scientific discoveries into clinical practice.

Journal Analysis

Journal analysis identified the core academic domains contributing to PKU research. The publications were distributed across 429 journals, of which 40 published at least 5 PKU-related articles. Figure 8 presents the journal co-citation network, visualizing the intellectual structure and disciplinary clustering of journals that have significantly contributed to PKU research. Table 5 summarizes the top 10 journals, including their Journal Citation Reports category, impact factor (IF), number of articles, total citations, publisher, and country of origin. *Molecular Genetics and Metabolism* ranked first in both the number of articles and total citations, reflecting its central role in disseminating research directly related to the genetic and metabolic aspects of PKU. Most high-frequency journals were associated with genetics, metabolism, pediatrics, nutrition, rare diseases, and endocrinology.

Figure 8. Journal co-citation network.



Although journals with moderate impact factors account for the majority of publications, several high-impact international journals have also featured PKU-related research. *The Lancet* (IF=98.4), *The New England Journal of Medicine* (IF=74.7), *Nature Medicine* (IF=58.7), *Cell Host & Microbe* (IF=30.3), and *Nature Communications* (IF=16.6) have each published 1 to 2 articles on PKU. Despite the small output, these publications exhibit citation frequencies far above the field average, highlighting both the visibility of PKU research and the importance of its broader clinical and scientific implications.

XSL•FO
RenderX

Table . Top 10 journals in publication.

Rank	Journal name	2023 IF[a]	JCR[b] category rank (quartile)	Publisher and country	Documents, n	Citations, n	Total link strength, n
1	*Molecular Genetics and Metabolism*	2.9	Q3 (Biochemistry & Molecular Biology)	Elsevier, United States	151	4506	1259
2	*Journal of Inherited Metabolic Disease*	4.6	Q2 (Genetics & Heredity)	Wiley, United States	116	3612	911
3	*Nutrients*	5.9	Q1 (Nutrition & Dietetics)	MDPI, Switzerland	76	570	620
4	*Molecular Genetics and Metabolism Reports*	1.9	Q4 (Biochemistry & Molecular Biology)	Elsevier, United States	56	759	536
5	*Orphanet Journal of Rare Diseases*	3.7	Q2 (Medicine, Research & Experimental)	Springer Nature, United Kingdom	32	560	334
6	*Journal of Human Nutrition and Dietetics*	3.3	Q2 (Nutrition & Dietetics)	Wiley, United Kingdom	17	273	115
7	*Journal of Pediatrics*	3.3	Q2 (Pediatrics)	Elsevier, United States	15	523	160
8	*Pediatrics*	5.8	Q1 (Pediatrics)	American Academy of Pediatrics, United States	15	470	123
9	*Journal of Pediatric Endocrinology & Metabolism*	1.8	Q4 (Pediatrics) or Q4 (Endocrinology & Metabolism)	Walter de Gruyter, Germany	14	91	60
10	*PLOS One*	3.2	Q2 (Multidisciplinary Sciences)	PLOS, United States	13	319	128

[a]IF: impact factor.

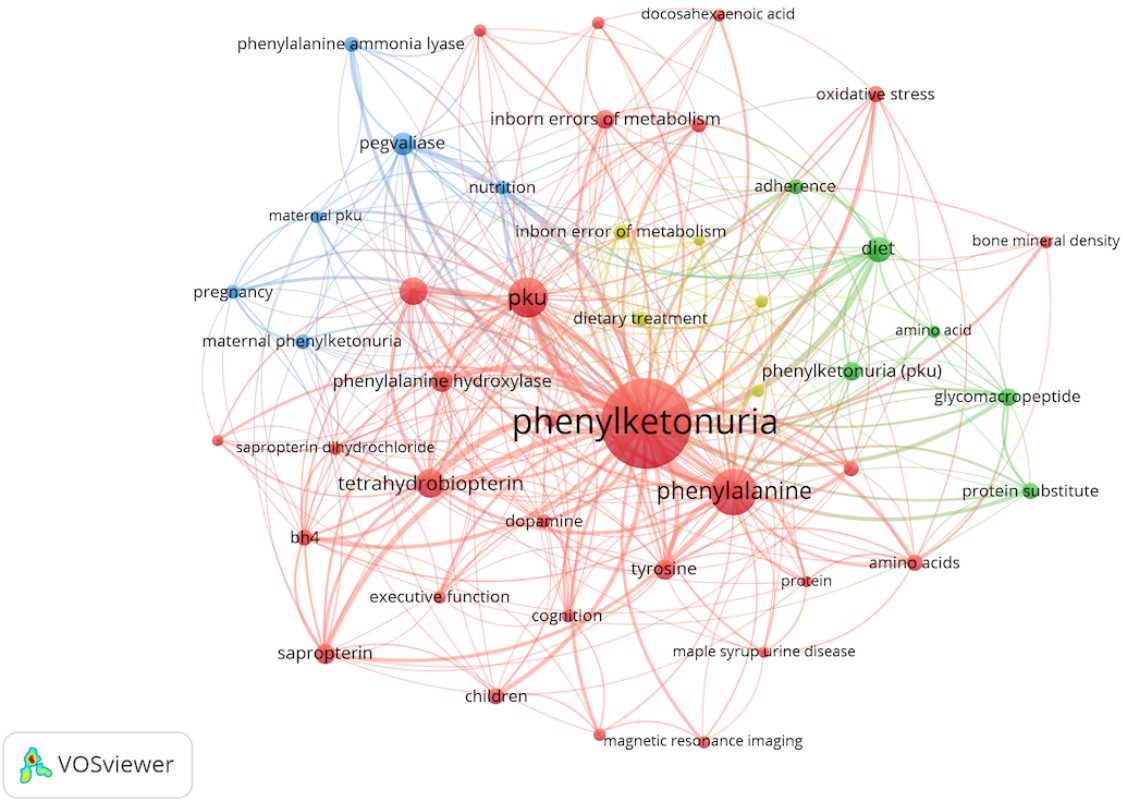
[b]JCR: Journal Citation Reports.

Keyword Analysis

A total of 2291 keywords were extracted from the included publications. Among these, 46 terms meeting the threshold of ≥10 occurrences were selected for co-occurrence network analysis. Figure 9 presents the resulting keyword co-occurrence network. In the visualization, node size represents keyword frequency, color gradient indicates the temporal distribution of research attention (cool colors for earlier periods, warm colors for recent hotspots), and edge thickness reflects co-occurrence strength.

XSL•FO
RenderX

Figure 9. Keyword co-occurrence network.



The results identified “phenylketonuria” (576 occurrences, centrality =0.74) as the dominant core term, forming a tightly connected research cluster with “phenylalanine metabolism” (156 occurrences), “tetrahydrobiopterin therapy” (60 occurrences), and “enzyme replacement therapy” (38 occurrences). In the domain of nutritional management, “dietary intervention” (46 occurrences), together with “protein substitutes” (19 occurrences) and “glycomacropeptide application” (23 occurrences), constituted the principal research axis. The strong association between “genotype analysis” (10 occurrences) and “hyperphenylalaninemia” (56 occurrences) highlights the increasing integration of precision medicine concepts into clinical practice. Notably, although terms such as “neurocognitive function” (13 occurrences) and “white matter integrity” (11 occurrences) appeared with moderate frequency, their relatively high betweenness centrality values (0.16 - 0.19) indicate that they act as critical bridging nodes, connecting biochemical research with clinical outcome studies and signaling their importance as emerging research foci.

Temporal evolution analysis revealed 3 distinct phases of shifting research priorities. In the early phase, between 2000 and 2009, research predominantly centered on “classic dietary therapy” and “newborn screening system optimization.” In the second phase, from 2010 to 2017, research transitioned toward precision medicine directions, including “tetrahydrobiopterin responsiveness” and “genotype-phenotype correlations.” The most recent phase, from 2018 to 2024, converged on 3 cutting-edge directions: “pegylated enzyme preparations” (burst strength 7.86), “metabolomics analysis” (strength 6.24), and “oxidative stress monitoring” (strength 5.93). These areas reflect a paradigm shift toward molecularly targeted therapies and dynamic metabolic monitoring approaches.

Figure 10 illustrates the top 25 keywords with the strongest citation bursts, which further demonstrate 3 distinct evolutionary phases. Analysis of these top 25 keywords with the strongest citation bursts further demonstrates 3 distinct evolutionary phases. The early phase (2000 - 2007) was defined by breakthroughs in biochemical mechanisms and cofactor therapy, as indicated by strong bursts for “tetrahydrobiopterin” (14.37) and “phenylalanine hydroxylase,” marking breakthroughs in cofactor therapy and metabolic pathway research. Concurrent bursts for “maternal PKU” (10.4) and “dietary treatment” highlighted foundational work in clinical management. The second phase (2011 - 2020) witnessed a shift toward neurocognitive and long-term outcome research. During this period, “oxidative stress” (6.25) revealed novel pathological mechanisms, while emerging terms such as “quality of life” (8.91) and “adult patients” marked the growing emphasis on holistic and lifespan-oriented management. The recent phase (2022 - 2025) exhibits precision management trends, with the keyword burst for “metabolic control” (5.12) highlighting increasing attention to individualized therapy, therapeutic optimization, and comprehensive outcome evaluation. Collectively, these trends reflect an overarching transition in PKU research from mechanistic exploration to clinical refinement, and ultimately toward precision and personalized medicine.

XSL•FO
RenderX

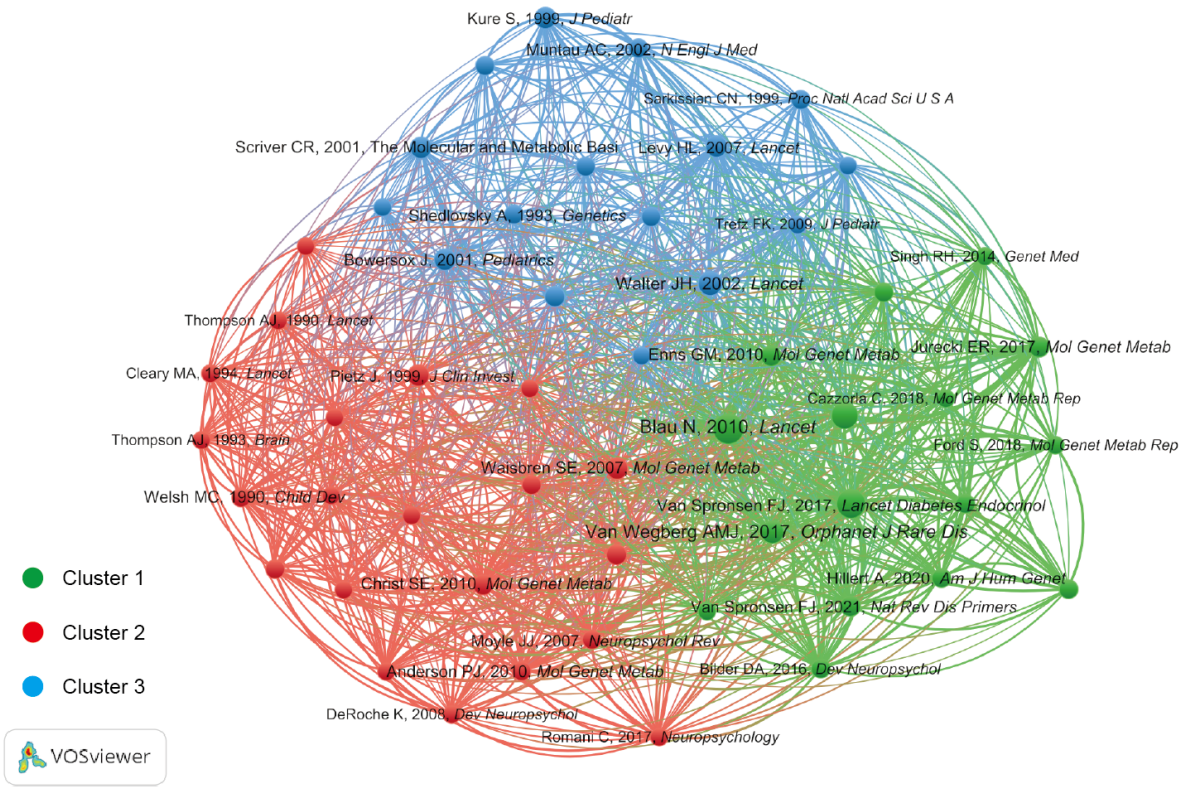
Figure 10. Top 25 keywords with the strongest citation bursts. The colors represent clusters automatically generated by VOSviewer. LHC: Large Hadron Collider.

Reference Co-Citation Analysis

Figure 11 displays the reference co-citation network of foundational literature in PKU research, with node size reflecting citation frequency and edge thickness denoting co-citation strength. The structure reveals several distinct yet interrelated knowledge domains that collectively constitute the intellectual foundation of PKU research. Additionally, Figure 12 presents the same co-citation network in the form of a heatmap, providing an alternative visualization of node importance and clustering intensity.

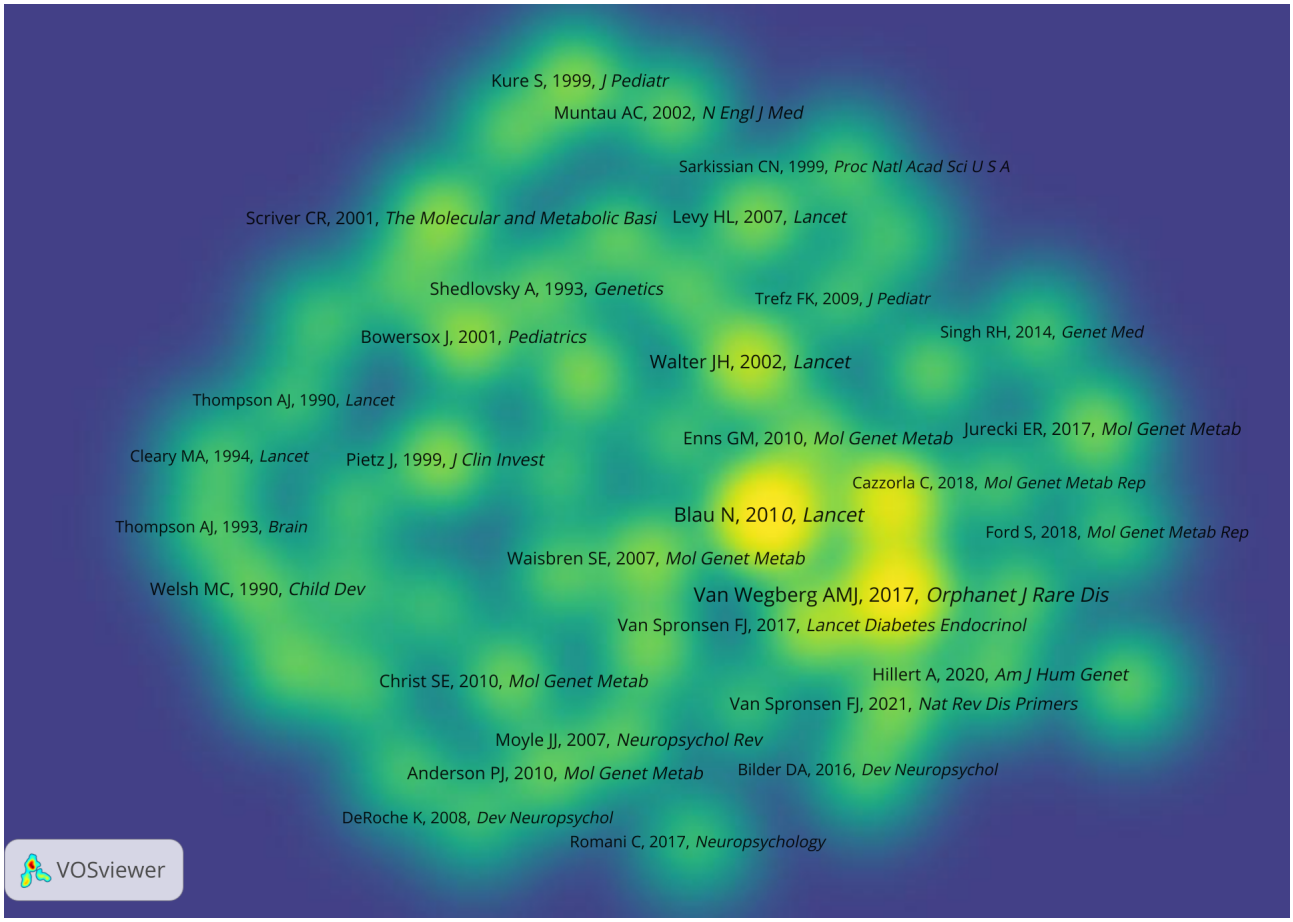
XSL•FO
RenderX

Figure 11. Co-citation of references [1,2,11-39]. The colors represent different clusters automatically generated by VOSviewer based on the co-occurrence analysis.



XSL•FO
RenderX

Figure 12. Hotspot map of co-cited references [1,2,11-39].



The most prominent node in the network is Blau et al [1] (*Lancet*, cited 233 times), which functions as a central integrative hub. This landmark publication on sapropterin responsiveness demonstrates strong co-citation links with both clinical management literature [11,12] and basic biochemical studies [17,19,40]. Its position indicates its significant role in bridging mechanistic insights with therapeutic applications.

Two major intellectual lineages emerge clearly from the network: one lineage, the clinical management tradition, connects early seminal works on PKU diagnosis and treatment [15,41] (with later authoritative guidelines and consensus statements [20]. Meanwhile, a second lineage, the basic science and mechanistic tradition, links foundational biochemical studies [17,42,43] with recent therapeutic advances [2].

The network also highlights strong co-citation clusters connecting neurocognitive outcome studies [21,22,44,45] with metabolic research [16,46]. This relationship reflects the long-standing scientific recognition that metabolic control is intimately linked to neurological integrity and cognitive development in PKU. Notably, historical milestones [23,47] maintain enduring co-citation ties with contemporary research, emphasizing the cumulative and evolving nature of knowledge construction in this field.

Overall, the co-citation patterns demonstrate a highly integrated research ecosystem, characterized by dynamic interplay between basic science discoveries and clinical innovation. This close coupling between bench and bedside has enabled rapid translation of mechanistic findings into therapeutic advances (Table 6).

XSL•FO
RenderX

Table . Top 10 most cited articles in publications.

Rank	Cited reference	Citations (n)	Total link strength (n)
1	Blau et al [1], 2010	233	884
2	van Wegberg et al [11], 2017	192	788
3	Vockley et al [48], 2014	150	746
4	Walter et al [15], 2002	112	480
5	van Spronsen et al [12], 2017	94	413
6	Bowersox [24], 2001	80	330
7	Pietz et al [16], 1999	78	287
8	Scriver and Kaufman [17], 2001	78	216
9	Waisbren et al [25], 2007	76	400
10	Kure et al [14], 1999	74	226

Discussion

Principal Findings

This bibliometric analysis provides a comprehensive overview of global PKU treatment research from 2000 to 2025. The findings reveal a sustained increase in publication output over the past 25 years, reflecting growing scientific and clinical interest in PKU management. The United States and several European countries remain the dominant contributors to the field, supported by extensive international collaboration networks. Keyword co-occurrence, citation burst, and thematic evolution analyses demonstrate a gradual transition from traditional dietary management toward pharmacological interventions, enzyme substitution therapies, neurocognitive outcome assessment, and precision medicine–oriented approaches. Collectively, these findings highlight the increasing diversification of PKU research and suggest that future investigations are likely to focus on individualized therapeutic strategies and emerging disease-modifying technologies.

General Information

Since 2000, global research on PKU has advanced substantially. The United States and several European countries, particularly the Netherlands, Germany, and the United Kingdom, have formed the central research axis in this field. These countries exhibit not only high publication productivity but also strong citation performance, reflecting their long-standing research foundations and academic leadership [1,11,12,48-51]. The current scientific landscape is distinctly internationalized, with dense collaborative links between North American and European institutions, as well as strong intra-European academic networks.

At the institutional and researcher levels, major contributions have largely originated from specialized metabolic centers and long-established clinical research groups. In the United States, researchers such as Cary O. Harding have been highly influential contributors to research on gene therapy and on pharmacological interventions in PKU and pharmacological innovation, with findings frequently published in high-impact journals, including *Molecular Genetics and Metabolism*, *The Lancet*, and *The American Journal of Clinical Nutrition* [52,53]. In the Netherlands, Francjan J. van Spronsen and Anita MacDonald have made seminal contributions in evidence-based dietary management, treatment optimization, and long-term outcome assessment, forming the empirical backbone for international PKU management guidelines [12,54]. Research teams in the United Kingdom further strengthened Europe's leadership through influential work on neurocognitive development, metabolic phenotyping, and patient-centered outcome evaluation.

Large-scale multinational collaborations, most notably the European PKU Guidelines Working Group, as well as other long-term outcome research consortia, have generated landmark findings that clearly depict the relationship between early metabolic control and neurodevelopmental outcomes, paving the way for individualized treatment strategies [11]. The extensive and highly interconnected co-authorship networks observed in this study reflect these sustained collaborative efforts, highlighting not only the central roles of Europe and the United States but also the emerging contributions from Asia and Australia. Overall, these global research activities have deepened scientific understanding of PKU pathophysiology, mechanisms, and treatment strategies and also laid a solid foundation for advancing precision medicine and novel therapies for individuals living with PKU worldwide.

Research Hotspots and Theme Evolution

Bibliometric analysis indicates that "metabolic control or nutritional management" constitutes the most central and persistently active research cluster in PKU, serving as the conceptual bridge linking major thematic domains over the past 2 decades. Converging evidence from both clinical and basic research supports a core consensus: achieving sustained and stable metabolic control is essential for optimizing long-term outcomes in patients with PKU [55]. Longitudinal cohort studies demonstrate that early-life fluctuations in blood phenylalanine levels are closely associated with later neurocognitive performance, while neuroimaging studies reveal that metabolic instability correlates strongly with disruptions in white matter microstructure [56]. Comprehensive systematic reviews, most notably the analysis by van Spronsen in *Molecular Genetics and Metabolism*, have validated a dose-response relationship between blood phenylalanine thresholds and neurological injury. Disturbances in metabolic homeostasis, particularly elevated

XSL•FO
RenderX

phenylalanine to tyrosine ratios, impair neurotransmitter synthesis and myelination, thereby heightening vulnerability to neurodevelopmental impairment [57]. Furthermore, phenotype-genotype association studies have linked PAH mutation profiles with metabolic phenotypes and treatment responsiveness, and Blau et al [1] have advanced a genotype-guided framework for individualized therapeutic decision-making.

A second major keyword cluster focuses on "neurocognitive and behavioral outcomes," encompassing concepts such as "executive function," "white matter integrity," and "cognitive development," reflecting the ongoing investigation into the mechanisms of long-term neurological impairment in PKU. Tyrosine deficiency results in compromised prefrontal cortex function and reduced cognitive flexibility. These insights have broadened the research perspective from traditional metabolic monitoring to investigations of neurochemical pathways and functional brain networks, reframing PKU as a modifiable neurometabolic disorder. Consequently, the clinical priority has evolved from determining whether intellectual disability occurs to how to optimize neurodevelopmental pathways through early metabolic control [58].

The third prominent research direction centers on *therapeutic strategies* and *individualized interventions*. Keywords such as "tetrahydrobiopterin responsiveness," "enzyme replacement therapy," "large neutral amino acids," and "glycomacropeptide" illustrate the transition from uniform dietary management to precision medicine [13,59-61]. Accumulating evidence supports the effectiveness of individualized therapeutic protocols tailored to genotype, metabolic phenotype, and pharmacological responsiveness, with demonstrated benefits for treatment adherence and quality of life. Current research priorities include defining optimal candidate populations for various treatment modalities (classic dietary therapy, cofactor treatment, and enzyme replacement therapy), developing models for predicting therapeutic efficacy, and refining protein substitute formulations to enhance metabolic stability. These research directions align with broader advancements in the management of inborn errors of metabolism and highlight the interdisciplinary integration of nutrition science, molecular biology, and pharmaceutical innovation [62,63].

From a chronological perspective, PKU research has progressed from macroscopic clinical observation to increasingly nuanced mechanistic understanding. During the early 2000s, investigations grounded in the "dietary control hypothesis" established treatment standards based on clinical observation and biochemical parameters. By approximately 2010, the rapid adoption of gene sequencing technologies propelled PAH genotype-phenotype correlations to the forefront, driving intensive examination of the metabolic consequences of specific mutations [64]. Since 2015, the emergence of long-term cohort studies, multimodal neuroimaging, and clinical trials of novel therapeutics has fostered the development of an integrated research paradigm that connects metabolic, genetic, neurocognitive, and behavioral dimensions of PKU [65,66]. Looking ahead, several areas hold strong potential for future breakthroughs: targeted modulation of key metabolic pathways to prevent neurotoxicity; integration of genetic backgrounds with metabolic phenotypes to identify high-risk subgroups and guide individualized interventions; and comprehensive characterization of the metabolic-brain-behavior axis in PKU through integrated models incorporating metabolomics, radiomics, and neurobehavioral profiling. Collectively, these advancements position precision metabolic management at the forefront of future PKU research and clinical practice [67,68].

Influential Literature and Research Paradigm Transformation

The early theoretical foundation of PKU research was shaped by the "metabolic imbalance hypothesis," which emerged in the late 20th century and was refined through the seminal work of Scriver [69]. Landmark cohort studies soon provided clinical validation for this hypothesis. Notably, investigations by Pietz et al [46] demonstrated a significant negative correlation between the duration of neonatal blood phenylalanine elevations above treatment thresholds and subsequent childhood IQ scores. These studies advanced the field beyond descriptive phenotypic characterization, establishing clear causal links between specific metabolic markers and neurodevelopmental outcomes, signaling a critical transition from macroscopic observation to mechanistic inquiry.

Around the year 2000, breakthroughs in molecular genetics and neuroimaging technologies catalyzed a profound methodological and conceptual transformation in PKU research. Using magnetic resonance spectroscopy, Sijens et al [70] provided the first direct evidence of metabolic alterations within cerebral white matter, challenging the long-standing assumption that neurological impairment was driven solely by circulating phenylalanine levels. Subsequent studies further delineated the association between cerebral metabolic patterns, executive function deficits, neurophysiological changes, and clinical severity, including analyses of large-scale brain network connectivity in adults with PKU. Currently, nutrition-focused research, exemplified by work from the European PKU Guideline Collaboration Group, integrates metabolomics and neuropsychological assessments to reveal how deficiencies in essential amino acids disrupt neurotransmitter synthesis and influence neurodevelopmental tendencies [71]. Together, these advances established "metabolic dysregulation affecting neurological function" as a core research paradigm and formed an interconnected knowledge framework that spans theoretical hypotheses, clinical observations, and mechanistic studies.

Findings from citation burst analysis indicate that the field is now undergoing a second paradigm shift, moving from mechanistic explanation toward clinical translation. Recent highly cited studies increasingly focus on the metabolic-brain axis, multi-omics integration, and precision nutrition strategies, reflecting the transformative impacts of high-throughput sequencing, metabolomics, and systems-level analytical platforms. With these technologies, researchers can characterize metabolic status in situ, map metabolic pathways, and model nutrient-metabolism-neural interactions in ways previously unattainable [62,71,72]. These advances are steering PKU research toward prevention-oriented metabolic management and precision medicine. Future priorities include the targeted restoration of key metabolic pathways to avert neurological

XSL•FO
RenderX

injury, the integration of genetic and metabolic profiles for risk stratification, and the application of multi-omics approaches to identify therapeutic targets. Collectively, the progression from empirical observation to mechanistic understanding, and finally to translational innovation, reflects a maturing research paradigm. It suggests that metabolism-guided strategies will be at the forefront of future PKU management. In addition to dietary management, pharmacological treatment, and enzyme substitution therapy, gene therapy has emerged as a promising frontier in PKU research. Recent advances in adeno-associated virus–mediated gene delivery, genome editing technologies, and liver-directed therapeutic strategies have demonstrated encouraging preclinical and early clinical results. Although challenges related to long-term safety, durability of therapeutic effects, and large-scale clinical implementation remain, gene therapy has the potential to provide sustained metabolic correction and may fundamentally transform the treatment paradigm for PKU in the future [9,10].

Limitations

This study has several limitations that should be acknowledged. First, the analysis was based exclusively on publications indexed in the WoSCC. Although WoSCC is widely recognized as a reliable source for bibliometric research, relevant studies indexed exclusively in other databases, such as PubMed, Scopus, or Embase, may have been omitted. Second, only English-language publications were included, which may have introduced language bias and resulted in the underrepresentation of research from non–English-speaking regions. Third, although bibliometric indicators such as publication counts and citation frequencies are useful for evaluating research productivity and academic influence, they do not directly reflect methodological quality, clinical significance, or scientific rigor. In addition, citation-based metrics may be influenced by factors such as self-citation and differences in citation practices across disciplines. Fourth, despite efforts to standardize author names, institutional affiliations, and country information, residual inconsistencies in database records may have affected certain collaboration and productivity analyses. Finally, because the literature search was conducted on September 13, 2025, the publication data for 2025 represent only a partial year and should therefore be interpreted with caution. Nevertheless, by integrating VOSviewer and CiteSpace and examining 25 years of scholarly output, this study provides a comprehensive overview of the evolution, collaboration patterns, and emerging research directions in PKU treatment research.

Conclusions

This bibliometric analysis systematically delineates the development of PKU research over the past 25 years. The field has advanced from an initial stage grounded in metabolic control and dietary intervention toward a mechanistic era characterized by elucidation of the metabolic-neural axis, and is now transitioning into a precision-medicine paradigm that integrates tailored nutrition, enzyme replacement therapies, and targeted metabolic regulation. Global research output continues to be dominated by the United States and several European countries, with China emerging as an increasingly influential contributor. Co-authorship networks and institutional collaboration patterns reveal a cohesive core of long-standing international collaborations that have consistently propelled high-impact progress. Keyword evolution and reference burst analyses jointly highlight individualized metabolic management through integrated multi-omics as the current frontier of PKU research.

Acknowledgments

OpenAI’s ChatGPT (GPT-5) was used solely for language refinement and editorial assistance. All scientific content, data interpretation, and conclusions were independently verified, critically reviewed, and approved by all authors. The use of ChatGPT did not influence the scientific validity or results of the study.

Funding

This work was supported by the CAMS Innovation Fund for Medical Sciences (2025-I2M-KJ018), the Fundamental Research Funds for the Central Universities, Peking Union Medical College (3332025164), the China National Postdoctoral Program for Innovative Talents (BX20240044), and the Lingang Laboratory Project (LGL-2615-07).

Data Availability

The datasets analyzed in this study were obtained from the Web of Science Core Collection, which is publicly available. The analysis code can be made available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: LW, KS, QC
Data curation: LW, SY, KS
Formal analysis: SY, QC
Investigation: KS, QC
Supervision: LW
Writing – original draft: SY
Writing – review & editing: LW, KS, QC, JJ

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Blau N, van Spronsen FJ, Levy HL. Phenylketonuria. Lancet 2010 Oct 23;376(9750):1417-1427. [doi: 10.1016/S0140-6736(10)60961-0] [Medline: 20971365]
2. van Spronsen FJ, Blau N, Harding C, Burlina A, Longo N, Bosch AM. Phenylketonuria. Nat Rev Dis Primers 2021 May 20;7(1):36. [doi: 10.1038/s41572-021-00267-0] [Medline: 34017006]
3. Appelberg K, Sörensen L, Zetterström RH, Henriksson M, Wedell A, Levin L. Cost-effectiveness of newborn screening for phenylketonuria and congenital hypothyroidism. J Pediatr 2023 May;256:38-43.e3. [doi: 10.1016/j.jpeds.2022.10.046] [Medline: 36495999]
4. Chen KY, Chan HC, Chan CM. How has global research on neural degeneration in glaucoma evolved from 1977 to 2025? A pioneering bibliometric insight into trends, impact, and future directions. Int Ophthalmol 2025 Aug 13;45(1):335. [doi: 10.1007/s10792-025-03700-5] [Medline: 40801966]
5. Chen KY, Chan HC, Chan CM. Global trends and developments in research on visually impaired athletes from 1960 to 2025: a bibliometric analysis. Eur J Ophthalmol 2026 Jul;36(4):997-1006. [doi: 10.1177/11206721261419580] [Medline: 41906345]
6. Ninkov A, Frank JR, Maggio LA. Bibliometrics: methods for studying academic publishing. Perspect Med Educ 2022 Jun;11(3):173-176. [doi: 10.1007/s40037-021-00695-4] [Medline: 34914027]
7. Arruda H, Silva ER, Lessa M, Proença D Jr, Bartholo R. VOSviewer and Bibliometrix. J Med Libr Assoc 2022 Jul 1;110(3):392-395. [doi: 10.5195/jmla.2022.1434] [Medline: 36589296]
8. Zhong D, Li Y, Huang Y, Hong X, Li J, Jin R. Molecular mechanisms of exercise on cancer: a bibliometrics study and visualization analysis via CiteSpace. Front Mol Biosci 2022;8:797902. [doi: 10.3389/fmolb.2021.797902] [Medline: 35096970]
9. Chen KY, Chan HC, Chan CM. Can gene therapy transform the treatment landscape of posterior segment eye diseases? A comprehensive review of recent advancements. Drugs 2025 Dec;85(12):1585-1608. [doi: 10.1007/s40265-025-02237-2] [Medline: 41138048]
10. Chen KY, Chan HC, Chan CM. Can stem cell therapy revolutionize ocular disease treatment? A critical review of preclinical and clinical advances. Stem Cell Rev Rep 2025 Jun;21(5):1160-1185. [doi: 10.1007/s12015-025-10884-x] [Medline: 40266467]
11. van Wegberg AMJ, MacDonald A, Ahring K, et al. The complete European guidelines on phenylketonuria: diagnosis and treatment. Orphanet J Rare Dis 2017 Oct 12;12(1):162. [doi: 10.1186/s13023-017-0685-2] [Medline: 29025426]
12. van Spronsen FJ, van Wegberg AM, Ahring K, et al. Key European guidelines for the diagnosis and management of patients with phenylketonuria. Lancet Diabetes Endocrinol 2017 Sep;5(9):743-756. [doi: 10.1016/S2213-8587(16)30320-5] [Medline: 28082082]
13. Hillert A, Anikster Y, Belanger-Quintana A, et al. The genetic landscape and epidemiology of phenylketonuria. Am J Hum Genet 2020 Aug 6;107(2):234-250. [doi: 10.1016/j.ajhg.2020.06.006] [Medline: 32668217]
14. Kure S, Hou DC, Ohura T, et al. Tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency. J Pediatr 1999 Sep;135(3):375-378. [doi: 10.1016/s0022-3476(99)70138-1] [Medline: 10484807]
15. Walter JH, White FJ, Hall SK, et al. How practical are recommendations for dietary control in phenylketonuria? Lancet 2002 Jul 6;360(9326):55-57. [doi: 10.1016/s0140-6736(02)09334-0] [Medline: 12114043]
16. Pietz J, Kreis R, Rupp A, et al. Large neutral amino acids block phenylalanine transport into brain tissue in patients with phenylketonuria. J Clin Invest 1999 Apr;103(8):1169-1178. [doi: 10.1172/JCI5017] [Medline: 10207169]
17. Scriver CR, Kaufman S. Hyperphenylalaninemia: phenylalanine hydroxylase deficiency. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW, et al, editors. The Metabolic and Molecular Bases of Inherited Disease, 8th edition: McGraw-Hill; 2001:1667-1724.
18. Trefz FK, Burton BK, Longo N, et al. Efficacy of sapropterin dihydrochloride in increasing phenylalanine tolerance in children with phenylketonuria: a phase III, randomized, double-blind, placebo-controlled study. J Pediatr 2009 May;154(5):700-707. [doi: 10.1016/j.jpeds.2008.11.040] [Medline: 19261295]
19. Sarkissian CN, Shao Z, Blain F, et al. A different approach to treatment of phenylketonuria: phenylalanine degradation with recombinant phenylalanine ammonia lyase. Proc Natl Acad Sci U S A 1999 Mar 2;96(5):2339-2344. [doi: 10.1073/pnas.96.5.2339] [Medline: 10051643]
20. Jurecki ER, Cederbaum S, Kopesky J, et al. Adherence to clinic recommendations among patients with phenylketonuria in the United States. Mol Genet Metab 2017 Mar;120(3):190-197. [doi: 10.1016/j.ymgme.2017.01.001] [Medline: 28162992]
21. Christ SE, Huijbregts SCJ, de Sonneville LMJ, White DA. Executive function in early-treated phenylketonuria: profile and underlying mechanisms. Mol Genet Metab 2010;99 Suppl 1:S22-S32. [doi: 10.1016/j.ymgme.2009.10.007] [Medline: 20123466]

XSL•FO
RenderX

22. Welsh MC, Pennington BF, Ozonoff S, Rouse B, McCabe ERB. Neuropsychology of early-treated phenylketonuria: specific executive function deficits. Child Dev 1990 Dec;61(6):1697-1713. [doi: 10.1111/j.1467-8624.1990.tb03560.x] [Medline: 2083493]
23. Muntau AC, Röschinger W, Habich M, et al. Tetrahydrobiopterin as an alternative treatment for mild phenylketonuria. N Engl J Med 2002 Dec 26;347(26):2122-2132. [doi: 10.1056/NEJMoa021654] [Medline: 12501224]
24. National Institutes of Health Consensus Development Panel, Bowersox J. National Institutes of Health Consensus Development Conference Statement: phenylketonuria: screening and management, October 16-18, 2000. Pediatrics 2001 Oct;108(4):972-982. [doi: 10.1542/peds.108.4.972] [Medline: 11581453]
25. Waisbren SE, Noel K, Fahrbach K, et al. Phenylalanine blood levels and clinical outcomes in phenylketonuria: a systematic literature review and meta-analysis. Mol Genet Metab 2007;92(1-2):63-70. [doi: 10.1016/j.ymgme.2007.05.006] [Medline: 17591452]
26. Levy HL, Milanowski A, Chakrapani A, et al. Efficacy of sapropterin dihydrochloride (tetrahydrobiopterin, 6R-BH4) for reduction of phenylalanine concentration in patients with phenylketonuria: a phase III randomised placebo-controlled study. The Lancet 2007 Aug;370(9586):504-510. [doi: 10.1016/S0140-6736(07)61234-3]
27. Shedlovsky A, McDonald JD, Symula D, Dove WF. Mouse models of human phenylketonuria. Genetics 1993 Aug;134(4):1205-1210. [doi: 10.1093/genetics/134.4.1205] [Medline: 8375656]
28. Singh RH, Rohr F, Frazier D, et al. Recommendations for the nutrition management of phenylalanine hydroxylase deficiency. Genet Med 2014 Feb;16(2):121-131. [doi: 10.1038/gim.2013.179] [Medline: 24385075]
29. Thompson AJ, Smith I, Brenton D, et al. Neurological deterioration in young adults with phenylketonuria. Lancet 1990 Sep 8;336(8715):602-605. [doi: 10.1016/0140-6736(90)93401-a] [Medline: 1975386]
30. Enns GM, Koch R, Brumm V, Blakely E, Suter R, Jurecki E. Suboptimal outcomes in patients with PKU treated early with diet alone: revisiting the evidence. Mol Genet Metab 2010;101(2-3):99-109. [doi: 10.1016/j.ymgme.2010.05.017] [Medline: 20678948]
31. Cleary MA, Walter JH, Wraith JE, et al. Magnetic resonance imaging of the brain in phenylketonuria. Lancet 1994 Jul 9;344(8915):87-90. [doi: 10.1016/s0140-6736(94)91281-5] [Medline: 7912392]
32. Cazzorla C, Bensi G, Biasucci G, et al. Living with phenylketonuria in adulthood: The PKU ATTITUDE study. Mol Genet Metab Rep 2018 Sep;16:39-45. [doi: 10.1016/j.ymgmr.2018.06.007] [Medline: 30069431]
33. Thompson AJ, Tillotson S, Smith I, Kendall B, Moore SG, Brenton DP. Brain MRI changes in phenylketonuria. Associations with dietary status. Brain 1993 Aug;116 (Pt 4)(4):811-821. [doi: 10.1093/brain/116.4.811] [Medline: 8353710]
34. Ford S, O'Driscoll M, MacDonald A. Living with Phenylketonuria: lessons from the PKU community. Mol Genet Metab Rep 2018 Dec;17:57-63. [doi: 10.1016/j.ymgmr.2018.10.002] [Medline: 30364670]
35. Moyle JJ, Fox AM, Arthur M, Bynevelt M, Burnett JR. Meta-analysis of neuropsychological symptoms of adolescents and adults with PKU. Neuropsychol Rev 2007 Jun;17(2):91-101. [doi: 10.1007/s11065-007-9021-2] [Medline: 17410469]
36. Anderson PJ, Leuzzi V. White matter pathology in phenylketonuria. Mol Genet Metab 2010;99 Suppl 1(suppl 1):S3-S9. [doi: 10.1016/j.ymgme.2009.10.005] [Medline: 20123467]
37. Bilder DA, Noel JK, Baker ER, et al. Systematic review and meta-analysis of neuropsychiatric symptoms and executive functioning in adults with phenylketonuria. Dev Neuropsychol 2016;41(4):245-260. [doi: 10.1080/87565641.2016.1243109] [Medline: 27805419]
38. DeRoche K, Welsh M. Twenty-five years of research on neurocognitive outcomes in early-treated phenylketonuria: intelligence and executive function. Dev Neuropsychol 2008;33(4):474-504. [doi: 10.1080/87565640802101482] [Medline: 18568900]
39. Romani C, Palermo L, MacDonald A, Limback E, Hall SK, Geberhiwot T. The impact of phenylalanine levels on cognitive outcomes in adults with phenylketonuria: effects across tasks and developmental stages. Neuropsychology 2017 Mar;31(3):242-254. [doi: 10.1037/neu0000336] [Medline: 28240926]
40. Burton BK, Grange DK, Milanowski A, et al. The response of patients with phenylketonuria and elevated serum phenylalanine to treatment with oral sapropterin dihydrochloride (6R-tetrahydrobiopterin): a phase II, multicentre, open-label, screening study. J Inherit Metab Dis 2007 Oct;30(5):700-707. [doi: 10.1007/s10545-007-0605-z] [Medline: 17846916]
41. Medical Research Council Working Party on Phenylketonuria. Recommendations on the dietary management of phenylketonuria. Report of Medical Research Council Working Party on Phenylketonuria. Arch Dis Child 1993 Mar;68(3):426-427. [doi: 10.1136/adc.68.3.426] [Medline: 8466250]
42. Scriver CR, Kaufman S, Eisensmith RC, Woo SLC. The hyperphenylalaninemias. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. The Metabolic and Molecular Bases of Inherited Disease, 7th edition: McGraw-Hill; 1995:1015-1075.
43. Surtees R, Blau N. The neurochemistry of phenylketonuria. Eur J Pediatr 2000 Oct;159 Suppl 2(S2):S109-S113. [doi: 10.1007/pl00014370] [Medline: 11043156]
44. Diamond A, Prevor MB, Callender G, Druin DP. Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. Monogr Soc Res Child Dev 1997;62(4):i-v. [doi: 10.2307/1166208] [Medline: 9421921]
45. Brumm VL, Azen C, Moats RA, et al. Neuropsychological outcome of subjects participating in the PKU adult collaborative study: a preliminary review. J Inherit Metab Dis 2004;27(5):549-566. [doi: 10.1023/b:boli.0000042985.02049.ff] [Medline: 15669671]

XSL•FO
RenderX

46. Pietz J, Benninger C, Schmidt H, Scheffner D, Bickel H. Long-term development of intelligence (IQ) and EEG in 34 children with phenylketonuria treated early. Eur J Pediatr 1988 May;147(4):361-367. [doi: 10.1007/BF00496411] [Medline: 3396592]
47. Lenke RR, Levy HL. Maternal phenylketonuria and hyperphenylalaninemia. N Engl J Med 1980 Nov 20;303(21):1202-1208. [doi: 10.1056/NEJM198011203032104] [Medline: 7421947]
48. Vockley J, Andersson HC, Antshel KM, et al. Phenylalanine hydroxylase deficiency: diagnosis and management guideline. Genet Med 2014 Feb;16(2):188-200. [doi: 10.1038/gim.2013.157] [Medline: 24385074]
49. MacDonald A, Gokmen-Ozel H, van Rijn M, Burgard P. The reality of dietary compliance in the management of phenylketonuria. J Inherit Metab Dis 2010 Dec;33(6):665-670. [doi: 10.1007/s10545-010-9073-y] [Medline: 20373144]
50. Smith WE, Berry SA, Bloom K, et al. Phenylalanine hydroxylase deficiency diagnosis and management: a 2023 evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). Genet Med 2025 Jan;27(1):101289. [doi: 10.1016/j.gim.2024.101289] [Medline: 39630157]
51. Thomas J, Levy H, Amato S, et al. Pegvaliase for the treatment of phenylketonuria: results of a long-term phase 3 clinical trial program (PRISM). Mol Genet Metab 2018 May;124(1):27-38. [doi: 10.1016/j.ymgme.2018.03.006] [Medline: 29653686]
52. Harding CO. Sepiapterin: a potential new therapy for phenylketonuria. Lancet 2024 Oct 5;404(10460):1284-1286. [doi: 10.1016/S0140-6736(24)01819-1] [Medline: 39368832]
53. Harding CO, Arnold G, Berry GT, et al. Phenylketonuria in adults: we know plenty, but there is much more to learn. Am J Clin Nutr 2025 Mar;121(3):741-743. [doi: 10.1016/j.ajcnut.2024.12.024] [Medline: 40044395]
54. Evers RAF, van Wegberg AMJ, MacDonald A, Huijbregts SCJ, Leuzzi V, van Spronsen FJ. Dietary liberalization in tetrahydrobiopterin-treated PKU patients: does it improve outcomes? Nutrients 2022 Sep 19;14(18):3874. [doi: 10.3390/nu14183874] [Medline: 36145250]
55. Vardy ERLC, MacDonald A, Ford S, Hofman DL. Phenylketonuria, co-morbidity, and ageing: a review. J Inherit Metab Dis 2020 Mar;43(2):167-178. [doi: 10.1002/jimd.12186] [Medline: 31675115]
56. Muri R, Maissen-Abgottspon S, Reed MB, et al. Compromised white matter is related to lower cognitive performance in adults with phenylketonuria. Brain Commun 2023;5(3):fcad155. [doi: 10.1093/braincomms/fcad155] [Medline: 37265600]
57. Blau N, MacDonald A, van Spronsen F. There is no doubt that the early identification of PKU and prompt and continuous intervention prevents mental retardation in most patients. Mol Genet Metab 2011;104:S1. [doi: 10.1016/j.ymgme.2011.10.007] [Medline: 22056112]
58. Leuzzi V, Chiarotti F, Nardecchia F, van Vliet D, van Spronsen FJ. Predictability and inconsistencies of cognitive outcome in patients with phenylketonuria and personalised therapy: the challenge for the future guidelines. J Med Genet 2020 Mar;57(3):145-150. [doi: 10.1136/jmedgenet-2019-106278] [Medline: 31484718]
59. Lammi C, Bollati C, Fiori L, et al. Glycomacropeptide (GMP) rescued the oxidative and inflammatory activity of free L-AAs in human Caco-2 cells: new insights that support GMP as a valid and health-promoting product for the dietary management of phenylketonuria (PKU) patients. Food Res Int 2023 Nov;173(Pt 1):113258. [doi: 10.1016/j.foodres.2023.113258] [Medline: 37803570]
60. Scala I, Riccio MP, Marino M, Bravaccio C, Parenti G, Strisciuglio P. Large neutral amino acids (LNAAs) supplementation improves neuropsychological performances in adult patients with phenylketonuria. Nutrients 2020 Apr 15;12(4):1092. [doi: 10.3390/nu12041092] [Medline: 32326614]
61. Scala I, Brodosi L, Gueraldi D, et al. Pegvaliase therapy for phenylketonuria: real-world case series and clinical insights. Mol Genet Metab 2024 May;142(1):108151. [doi: 10.1016/j.ymgme.2024.108151] [Medline: 38522180]
62. Verduci E, Carbone MT, Borghi E, Ottaviano E, Burlina A, Biasucci G. Nutrition, microbiota and role of gut-brain axis in subjects with phenylketonuria (PKU): a review. Nutrients 2020 Oct 29;12(11):3319. [doi: 10.3390/nu12113319] [Medline: 33138040]
63. Ubaldi F, Frangella C, Volpini V, Fortugno P, Valeriani F, Romano Spica V. Systematic review and meta-analysis of dietary interventions and microbiome in phenylketonuria. Int J Mol Sci 2023 Dec 13;24(24):17428. [doi: 10.3390/ijms242417428] [Medline: 38139256]
64. Bashyam MD, Chaudhary AK, Reddy EC, et al. Phenylalanine hydroxylase gene mutations in phenylketonuria patients from India: identification of novel mutations that affect PAH RNA. Mol Genet Metab 2010 May;100(1):96-99. [doi: 10.1016/j.ymgme.2010.01.016] [Medline: 20188615]
65. Muri R, Rummel C, McKinley R, et al. Transient brain structure changes after high phenylalanine exposure in adults with phenylketonuria. Brain 2024 Nov 4;147(11):3863-3873. [doi: 10.1093/brain/awae139] [Medline: 38723047]
66. Longo N, Siriwardena K, Feigenbaum A, et al. Long-term developmental progression in infants and young children taking sapropterin for phenylketonuria: a two-year analysis of safety and efficacy. Genet Med 2015 May;17(5):365-373. [doi: 10.1038/gim.2014.109] [Medline: 25232857]
67. Garrett R, Ptolemy AS, Pickett S, Kellogg MD, Peake RWA. Untargeted metabolomics for inborn errors of metabolism: development and evaluation of a sustainable reference material for correcting inter-batch variability. Clin Chem 2024 Dec 2;70(12):1452-1462. [doi: 10.1093/clinchem/hvae141] [Medline: 39365746]
68. Koppes EA, Redel BK, Johnson MA, et al. A porcine model of phenylketonuria generated by CRISPR/Cas9 genome editing. JCI Insight 2020 Oct 15;5(20):e141523. [doi: 10.1172/jci.insight.141523] [Medline: 33055427]

XSL•FO
RenderX

69. Scriver CR. The PAH gene, phenylketonuria, and a paradigm shift. Hum Mutat 2007 Sep;28(9):831-845. [doi: 10.1002/humu.20526] [Medline: 17443661]
70. Sijens PE, Oudkerk M, Reijngoud DJ, Leenders KL, de Valk HW, van Spronsen FJ. 1H MR chemical shift imaging detection of phenylalanine in patients suffering from phenylketonuria (PKU). Eur Radiol 2004 Oct;14(10):1895-1900. [doi: 10.1007/s00330-004-2388-z] [Medline: 15378255]
71. Parolisi S, Montanari C, Borghi E, et al. Possible role of tryptophan metabolism along the microbiota-gut-brain axis on cognitive & behavioral aspects in phenylketonuria. Pharmacol Res 2023 Nov;197:106952. [doi: 10.1016/j.phrs.2023.106952] [Medline: 37804926]
72. van der Goot E, Vink SN, van Vliet D, van Spronsen FJ, Falcao Salles J, van der Zee EA. Gut-microbiome composition in response to phenylketonuria depends on dietary phenylalanine in BTBR Pahenu2 mice. Front Nutr 2022;8:735366. [doi: 10.3389/fnut.2021.735366] [Medline: 35059423]

Abbreviations

IF: impact factor
PAH: phenylalanine hydroxylase
PKU: phenylketonuria
WoSCC: Web of Science Core Collection

Edited by K Gierend; submitted 03.Feb.2026; peer-reviewed by CM Chan, F Jiang, G Wang, X Zhang; revised version received 08.Jun.2026; accepted 11.Jun.2026; published 03.Sep.2026.

Please cite as:
Yang S, Song K, Chen Q, Jiang J, Wang L
Global Trends in Phenylketonuria Treatment Research, 2000-2025: Bibliometric Analysis
Online J Public Health Inform 2026;18:e90419
URL: https://ojphi.jmir.org/2026/1/e90419
doi:10.2196/90419

© Sitong Yang, Kaichao Song, Qingbo Chen, Jiandong Jiang, Lulu Wang. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 3.Sep.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Comprehensive Approach to Days' Supply Estimation in a Real-World Prescription Database: Algorithm Development and Validation Study

Maria Malk[1], MSc; Kerli Mooses[1], PhD; Marek Oja[1], PhD; Johannes Holm[1], MSc; Hanna Keidong[1], MSc; Nikita Umov[2], MD; Sirli Tamm[1], MSc; Sulev Reisberg[1], PhD; Jaak Vilo[1], PhD; Raivo Kolde[1], PhD

[1]Institute of Computer Science, University of Tartu, Tartu, Tartu, Estonia

[2]Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia

Corresponding Author:
Maria Malk, MSc
Institute of Computer Science
University of Tartu
Narva mnt 18
Tartu, Tartu, 51009
Estonia
Phone: 372 56689471
Email: maria.malk@ut.ee

Abstract

Background: For accurate medication usage statistics and medication adherence calculations, we need to have an accurate days' supply (DS) for each prescription. Unfortunately, often the DS or the information needed for calculating the DS is not provided. Therefore, other methods need to be applied to acquire missing values or substitute incorrect values.

Objective: This study aims to apply a variety of methods for managing incomplete and missing data to enhance the accuracy of calculating DS for all medications and drug forms alike. Furthermore, to describe the effect of applied methods on the medication adherence calculated on real-world data.

Methods: A dataset comprising prescription records from a 10% (150,824 patients) random sample of the Estonian population between 2012 and 2019 was used. The workflow consisted of 3 steps: data cleaning, imputation, and calculation of DS. For imputation, different methods were combined, such as calculating mode-based daily dose, or using usage guidelines from the Summary of Product Characteristics or legislation. DS was calculated based on the provided daily dose or imputed value. To evaluate the impact of data cleaning, medication adherence for the baseline dataset and corrected dataset for 2 time periods, 2012-2015 and 2017-2019, was calculated and compared.

Results: The drug forms with the lowest proportion of correct DS provided were insulin injections (2601/82,867, 3.1%) and intravaginal contraceptives (1692/21,145, 8%) while the highest proportion of DS was provided for inhalation medication (78,541/126,588, 62%), oral drops (52,085/98,221, 53%) and tablets, capsules, suppositories (2,828,617/6,176,585, 45.8%). As a result of applying different imputation approaches, we successfully found the DS for 98.3% (7,415,347/7,544,892) of dispensed prescriptions. For the remaining 1.7% (129,545/7,544,892) of prescriptions, DS could not be imputed nor calculated with these methods. As for the medication adherence, the distinction between 2 observed time periods was more distinct in the baseline dataset compared with the corrected dataset for most of the drug groups, indicating that the applied correction methods had lessened the stark contrast.

Conclusions: In summary, our study demonstrated that with a carefully designed imputation pipeline where data-driven imputation is combined with domain knowledge and literature information, it is possible to meaningfully improve the quality of prescription datasets and generate more accurate and consistent adherence metrics across various drug forms. Nonetheless, future efforts should continue to refine imputation techniques, incorporate machine learning approaches where appropriate, and expand validation efforts using external benchmarks or clinical outcomes.

(Online J Public Health Inform 2026;18:e83465) doi:10.2196/83465

XSL•FO
RenderX

KEYWORDS

prescription records; days' supply; medication adherence; CMA; continuous multiple interval measures of medication availability; daily dose; imputation

Introduction

Electronic health care databases provide a valuable data source for conducting various studies, as they are detailed, structured, and often cover a long time span. One important source of such data is pharmacy medication records, which provide the opportunity to research medication usage and adherence cost-effectively and at scale [1-3]. To calculate medication adherence, an accurate days' supply (DS) for each prescription is needed [4]. DS describes how many days the dispensed medication is expected to last. In medical fields, DS is also referred to as treatment course length. While some prescription databases have DS recorded, in others, the DS value needs to be calculated using other available information, such as the number of dispensed medications and prescribed daily dose [4]. Unfortunately, there is an abundance of evidence suggesting inaccuracies and missing values in the prescription databases [4,5].

Several studies have specifically addressed the challenges related to missing daily doses and the DS issue [4,6,7]. For example, imputing 1 dose of medication per day for missing daily doses has been shown to work for patients with stroke [6], while imputing the mode daily dose per active substance and number of tablets per prescription has been shown to work for diabetic drugs [4]. Studies have also demonstrated that daily dose values can be imputed using machine learning (ML) algorithms that incorporate various patient characteristics [4]. Some studies [8,9] have applied the defined daily dose (DDD) toolkit developed by the World Health Organization [10]. However, it has been concluded that using DDD as a daily dose substitute may lead to misclassification of medication adherence [8,9]. The limitations of the existing studies tackling the missing daily dose and DS issue are that they often focus on 1 disease or medication group [4,6,7,9,11-13] and thus, it is unknown whether the same approach applies to other active substances or diseases. Moreover, studies have often been conducted using single-dose oral medications [4,6,7,9], and little is known about how to address the missing data in other drug forms such as eye drops, topical creams, and gels. Although some studies have researched medication adherence among diseases that often use other drug forms [11,13-18], only a few of these use pharmacy records [13,17]. However, these studies have not tackled the problem of missing data.

In addition to missing data, some studies have highlighted that some inaccuracies may be present in DS values [5,19]. More common inconsistencies are in reported DS values, dosage, fill intervals, administration times and quantity [5,19]. It has been stressed that further research is needed to evaluate DS reporting errors and to recommend strategies to address these errors [20].

To the best of our knowledge, no comprehensive approach exists that addresses both missing and inaccurate data in prescriptions database across all prescriptions, irrespective of the active substance or drug form. To address this shortcoming, this study aims to apply a variety of methods for managing incomplete and missing data to enhance the accuracy of calculating DS for all medications and drug forms alike. In addition, we describe the effect of applied methods on the medication adherence calculated on real-world data.

Methods

Data

The dataset used in this study consisted of prescription data for a 10% (150,824 patients) random sample of the Estonian population, covering the period from 2012 to 2019 [21]. The data originated from the national e-prescription database, where all prescriptions issued in primary and secondary care have been stored since 2010. The dataset includes all prescribed medications together with their dispensing information. Specific information about the prescriptions is shown in Table 1. Notably, the dataset does not contain information about over-the-counter medications or inpatient medications.

XSL•FO
RenderX

Table 1. Available prescription information.

Categories	Details
Patient information	• Unique ID for each patient
General administrative information	• Unique ID for each prescription • Issue date • Validity date • Dispensed date • Prescription type (initial, refill, narcotic, and medical device) • Prescribing health care provider information • Diagnosis code for which the medication was prescribed • Rationale for brand-specific prescribing • Dispensing pharmacy information • Prescription cancellation reason
Active substance	• Names of active substances • Amount of maximum 3 components • Unique ATC[a] code for active substance
Drug form	• Type of medication (eg, tablet, cream, and eye drops)
Daily dose information	• Number of times the medication is taken or appliedb • Amount of medication per doseb • Optional free-text
Days' supply	• Optional treatment course length
Dispensed package information	• Unique ID for each drug package • Number of packages • Number of units in each package • Item size (for non-single-dose medications, eg, creams and eye drops)

[a]ATC: Anatomical Therapeutic Chemical.

These multiplied make up the daily dose.

In total, the dataset initially comprised 9,279,082 prescriptions, of which 7,544,892 (81.3%) were dispensed to patients. The rest of the prescriptions were prescribed but never dispensed. As it stands, only the prescriptions that were dispensed were included in this study and equipped with a DS value.

There were several data quality issues in the database that had to be addressed. Providing daily dose information became mandatory for doctors in mid-2016. As a result, a 67.4% (3,070,512/4,555,074) of prescriptions issued prior to this date lack daily dose information compared with 0.4% (12,468/2,989,805) of prescriptions issued in 2017-2019. Throughout this study period, it was optional for the doctors to specify the treatment course length for each prescription. This treatment course length is equivalent to the DS and could be used as a substitute or for comparison with the calculated DS. The number of prescriptions with provided DS increased over time–in 2012-2016, 15.2% (693,868/4,555,074) of prescriptions had this information compared with 36.1% (1,078,710/2,989,805) in 2017-2019, calculated in the respective timeframes.

Even though almost all prescriptions after 2016 were provided with a daily dose, the information was inconsistent in terms of the units used. The amount of medication could be given as a quantity (eg, 1 tablet, 2 pills, etc) or the amount of active substance (eg, 10 mg, 20 mg, etc). Neither of these writing methods was consistent among themselves either. For example, if a drug contained 2 active substances, like 5 mg + 10 mg, and the doctor decided to write the amount taken at once in active substance amount, they had 3 distinct ways to write that: 5 mg, 10 mg, and 15 mg.

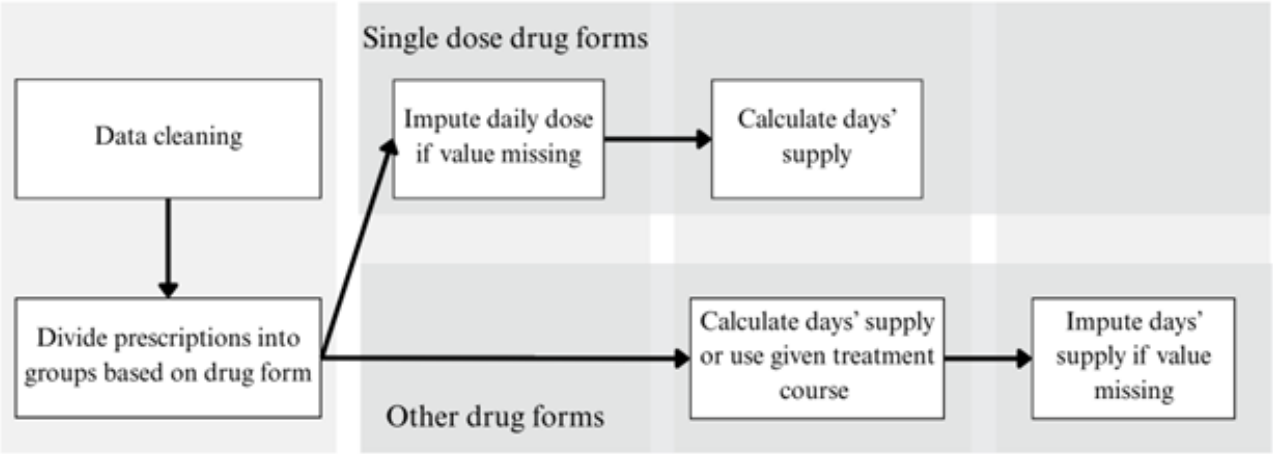
For non–single-dose medications such as drops, creams, and gels, the daily dose was often noted only qualitatively (eg, "once per day"), without quantitative detail, making DS impossible to calculate.

Workflow

The main workflow to acquire the DS value is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Workflow to acquire days' supply.



Data Cleaning and Drug Form Grouping

The prescriptions were first checked for correctness. Some of the prescriptions from 2012 had false active substance data. Since every dispensed medication has a package number, the correct active substance was taken from the package information.

Prescriptions were stratified by drug form into distinct groups, since each category follows specific conventions for stating the daily dose. For example, liquid formulations record daily doses in drops, creams in standardized dose units, and sprays in actuations. If needed, these groups were further divided by active substances because DS can be imputed more correctly for pharmacologically similar active substances and drug forms. For example, ophthalmic preparations were initially divided by drug form into 2 groups: single-dose containers and multiple-dose containers. The multiple-dose container group was further subdivided into 2 categories: one containing anti-infective, anti-inflammatory agents, and their combinations, and another containing all other substances. The last 2 subgroups required different imputation values.

All tablets, capsules, and suppositories share similar daily dose parameters, which allows for a consistent approach in calculating DS. These were collectively referred to as single-dose medications. Groups containing fewer than 1000 prescriptions were given a DS value based on the provided DS data and were excluded from further imputation processes.

Single-Dose Drug Forms DS

For single-dose medications, if all the necessary daily dose information, consisting of the amount of medication in a dose and the number of doses in a time unit (ie, day or week), was available, we calculated the DS. Since there are many different ways to prescribe the medication amount in 1 dose, we used a method that took into consideration whether the amount of medication in a dose was written in units (1 tablet, 2 tablets, etc) or active substance amount (5 mg, 10 mg, etc). The number of doses per time unit, which is required for daily dose estimation, was generally accurate and therefore used without modification. Afterwards, we compared the calculated value with the provided DS, if it was available. Whichever of the 2—calculated or provided DS—had a smaller value, it was used. The rationale is that a pharmacist may dispense a larger package (ie, more tablets) if no package exactly matches the prescription. If the provided DS was not available, the calculated DS was used.

For prescriptions where DS could not be determined using the aforementioned logic, a mode-based imputation method was used. A reference dataset was created with the most common daily dose for each active substance and corresponding dose strength. When 2 or more daily dose values were equally frequent, those active substances were excluded from imputation, and their prescriptions remained unimputed. DS was then calculated for imputed prescriptions. This approach adds dose specificity to the active substance classification, making the DS value more reliable compared to direct DS imputation.

Other Drug Forms DS

The general workflow for determining DS in the other dosage - form groups was as follows. The target drug group was thoroughly analyzed to identify trends in how daily doses are recorded and the specific characteristics of its treatment regimens. Where possible, the DS was calculated, and the existence of the prescribed DS was checked. If prescribed DS information was available, it was either used directly or compared to the computed DS value, and the more appropriate of the 2 was adopted. This comparison was consistently applied and therefore not explicitly restated in the subsequent list.

In cases where neither of these approaches could be applied, alternative imputation strategies were used. These strategies were based on the Summary of Product Characteristics (SPCs) [22], where the recommended DS is written and on relevant Estonian regulations [23], where, for example, the recommended prescription usage length is recorded. These outcomes were then reviewed by a domain expert and, if necessary, adjusted to reflect current practices in the field. Based on this evaluation, specific imputation rules were defined for each medication group and subsequently applied in the processing workflow.

More specifically, the DS was calculated or imputed for different drug forms as follows:

1. For semisolid medications consisting of creams, gels, shampoos, and ointments for topical usage, daily dosing

XSL•FO
RenderX

instructions are often provided without quantitative details. Consequently, implementing a DS calculation method would give limited value and was therefore not applied. If the provided DS was available, that value was used. Otherwise, it was imputed as 30 days per package.

2. For medicinal nail polish, daily dosing instructions are often provided without quantitative details. Therefore, DS calculations were not done. When the provided DS was available, it was used. Otherwise, DS was imputed based on the package type. As only 2 package types were present in the dataset, the DS was either 210 days or 180 days.
3. For eye drops, daily dosing instructions are often provided without quantitative details. Therefore, the DS calculation was not done. If the provided DS was available, it was used. Otherwise, due to shelf-life limits, DS was imputed as 30 days per package, except for anti-infectious, anti-inflammatory, and corticosteroid drops, which should not exceed a 2-week treatment course. Therefore, DS were imputed with either 30 days or 14 days, depending on the active substance. When the eye drops were packaged in a single-dose container, and a daily dose was provided, the daily dose was used to calculate DS. If daily data were not available, the provided DS was used, and if neither was available, it was imputed as 30 days.
4. For ear drops, all medications present in the dataset were for short-term use only. Namely, anti-infective and analgesic medications, which are usually not used for more than 7 days without a doctor's supervision. Since the daily dosing instructions are often provided without quantitative details, the DS calculations were not done. When the provided DS was available, it was used. Otherwise, a 7-day value was used for imputation.
5. For oral drops, if the daily dose was specified quantitatively (eg, in number of drops or amount of active substance), this information was used to calculate the DS. The calculations varied depending on the active substance, as the drop sizes differed between substances. When the DS could not be calculated, and provided DS was available, it was used. Otherwise, 30 days per package was imputed due to shelf-life limits.
6. For nasal sprays, DS was calculated based on the daily dose if it was provided. If the DS could not be calculated, we used the provided DS if it was available. Otherwise, a default imputed value of 30 days per package was applied except for the antifungal nasal spray, where the treatment course should not last more than 7 days. Therefore, for imputation, the 7-day value was used.
7. For inhalation powders, if the medication was divided into blisters and a daily dose was provided, then it was used to calculate DS. When the DS could not be calculated, and provided DS was available, it was used. Otherwise, a value of 60 days was imputed for all prescriptions, considering the prescription guidelines and prevailing trends. According to Estonian legislation and prescribing practices, chronic medications are issued on a single prescription for a 2-month period.
8. For syrups, if the daily dose was specified quantitatively, it was used to calculate the DS. When the DS could not be calculated, and provided DS was available, it was used. Otherwise, 30 days per package was imputed due to shelf-life limits.
9. For antibiotic solutions, daily dosing instructions are often provided without quantitative details, therefore DS calculation was not done. When the provided DS was available, it was used. For these medications, the treatment course should not be longer than 14 days; therefore, for imputation 14-day value was used.
10. For transdermal patches, daily dosing instructions are often provided without quantitative details, therefore DS calculation was not done. When the provided DS was available, it was used. Otherwise, the imputable DS was calculated based on the length of the effect of the patches based on information from SPCs.
11. For vaccines and implants, the treatment course was imputed as 1 day. Since there is no standard way for doctors to write a DS value, it was not used.
12. For insulin, a value of 60 days was imputed for all prescriptions, considering the prescription guidelines and prevailing trends. According to Estonian legislation and prescribing practices, chronic medications are issued on a single prescription for a 2-month period.

Quality Control Using Medication Adherence

To assess the impact of data cleaning and imputation, we calculated the mean value of DS together with 95% CI values for 147 active substances in the corrected datasets used in chronic conditions for 2 time periods: 2012-2015, when providing daily dose information was voluntary and 2017-2019, when this information was mandatory for the doctors. Since 2016 was a transition year, it was excluded from the time-period comparison. This approach shows us whether imputation introduced systematic bias in prescription duration. Moreover, we calculated medication adherence for these active substances in both the baseline and corrected datasets. This approach provides a broader statistical perspective across all medications, allowing us to evaluate the validity of the imputations based on all the prescriptions available.

The active substances used in the analysis were selected as follows: 300 of the most frequently prescribed active substances were extracted from the database, and 2 pharmacists independently filtered out those that were meant for chronic conditions. The active substances used in calculations were divided into 27 groups based on the Anatomical Therapeutic Chemical (ATC) therapeutic subgroup (Multimedia Appendix 1).

In the baseline dataset, the DS was calculated based on the provided daily dose for single-dose medications. When the DS was not provided, a value of 30 days was imputed [4]. Baseline dataset analyses were performed using a simplified approach, which is fast, inexpensive, and requires minimal data processing. In the corrected dataset, we applied more resource-intensive methods to evaluate whether such refinement provides a meaningful advantage in adherence estimation.

For medication adherence calculations, the continuous multiple interval measures of medication availability (CMA) [2] were used. Out of 8 CMAs, CMA5 was selected as it accounts for gaps in medication availability and assumes that the new refill

XSL•FO
RenderX

is stored until the previous prescription is fully used. The adherence was calculated on a yearly basis. The calculation window began with the first medication dispensing and ended with the last, requiring each patient to have at least 2 prescriptions dispensed. Any unused medication remaining at the end of the window was excluded from the calculations [2]. The CMA implementation in AdhereR [24] was used through AdherenceFromOMOP [25].

To describe the change in medication adherence between 2 periods (2012-2015 vs 2016-2019), the change in period means was calculated. Although the year 2016 was a transition year and excluded from the time-period comparison, the data for 2016 are presented in CMA figures in Multimedia Appendix 1.

Ethical Considerations

The study was approved by the Research Ethics Committee of the University of Tartu (300/T-23), the Estonian Committee on Bioethics and Human Research (1.1-12/653), and the requirement for informed consent was waived.

Results

A total of 7,544,892 dispensed prescriptions were included in the process of establishing the DS value. The largest drug form group was single-dose medications, including tablets, capsules, and suppositories, which accounted for 81.9% (n=6,176,585/7,544,892) of all dispensed prescriptions. The remaining drug form groups are listed in Table 2. In total, 13 major drug form categories were identified.

XSL•FO
RenderX

Table 2. Prescription distribution by drug groups and days' supply establishing methods, for the full dataset (2012-2019).

Drug group and dataset information	Method for finding days, supply	Prescriptions in main group, n (%)
Tablets, capsules, and suppositories		
Daily dose provided	Calculated using the given daily dose	2,828,617 (45.80)
Days' supply provided	Days' supply was used	933,710 (15.12)
Neither daily dose nor days' supply provided	Calculated using the imputed daily dose value	2,402,392 (38.89)
Days' supply could neither be calculated nor imputed	__[a]	11,866 (0.19)
Semisolid dosage forms		
Days' supply provided	Days' supply was used	119,748 (37.55)
Days' supply not provided	Imputed as 30 days per package[b]	199,154 (62.45)
Medicinal nail polish		
Days' supply provided	Days' supply was used	3008 (41.55)
Days' supply not provided	Imputed as 210 days or 180 days[b]	4231 (58.45)
Ear drops		
Days' supply provided	Days' supply was used	5661 (57.11)
Days' supply not provided	Imputed as 7 days per prescription[b]	4252 (42.89)
Eye drops		
Eye drops in single-dose containers - daily dose provided	Calculated using the given daily dose	6724 (2.35)
Other eye drops - days' supply provided	Days' supply was used	62,332 (21.81)
Other eye drops - days' supply not provided	Imputed as 14[b] or 30[c] days per package	216,740 (75.82)
Days' supply not calculated nor imputed	—	43 (0.02)
Oral drops		
Daily dose provided	Calculated using the given daily dose	52,085 (53.03)
Days' supply provided	Days' supply was used	33,055 (33.65)
Neither daily dose nor days' supply provided	Imputed as 30 days per prescription[b]	12,961 (13.20)
Days' supply neither calculated nor imputed	—	120 (0.12)
Inhalation medication		
Daily dose provided	Calculated using the given daily dose	78,541 (62.04)
Days' supply provided	Days' supply was used	19,885 (15.71)
Days' supply not provided	Imputed as 60 days per prescription[b]	27,137 (21.44)
Days' supply not calculated nor imputed	—	1025 (0.81)
Nasal sprays		
Antifungal nasal spray - days' supply not provided	Imputed as 7 days per prescription[b]	7738 (8.72)
Daily dose provided	Calculated using the given daily dose	26,361 (29.70)
Days' supply provided	Days' supply was used	40,551 (45.69)
Neither daily dose nor days' supply provided	Imputed as 30 days per prescription[b]	14,094 (15.88)
Syrups		
Antibiotics - days' supply not provided	Imputed as 14 days per prescription[b]	29,130 (34.31)
Daily dose provided	Calculated using the given daily dose	37,251 (43.87)
Days' supply provided	Days' supply was used	3094 (3.64)
Days' supply not provided	Imputed as 30 days per package[c]	15,436 (18.18)
Transdermal patch		

XSL•FO
RenderX

Drug group and dataset information	Method for finding days, supply	Prescriptions in main group, n (%)
Hormonal patch - days' supply not provided	Impute as 7 days per patch plus 7 days[b]	13,131 (81.18)
Analgesic patch - days' supply not provided	Imputed as 3 days per patch[b]	1127 (7.01)
Days' supply provided	Days' supply was used	599 (5.94)
Days' supply not provided	Impute as 7 days per patch[b]	865 (5.38)
Intravaginal contraceptives		
Days' supply provided	Days' supply was used	1692 (8)
Days' supply not provided	Imputed as 30 days per item in package[b]	19,453 (92)
Implants and vaccines		
Days' supply provided	Days' supply was used	18,940 (45.2)
Days' supply not provided	Imputed as 1 day	22,979 (54.8)
Insulin injections		
Days' supply provided	Days' supply was used	2601 (3.14)
Days' supply not provided	Imputed as 60 days per prescription[b]	80,266 (96.86)
Others		
Days' supply provided	Days' supply was used	69,806 (37.47)
Days' supply not calculated nor imputed	—	116,491 (62.52)

[a]Not available.

[b]Based on the Summary of Product Characteristics issued by the Estonian Agency of Medicines.

[c]Medication not shelf-stable for more than 30 days after opening.

The drug forms with the lowest proportion of correct DS provided were insulin injections (2601/82,867, 3.1%) and intravaginal contraceptives (1692/21,145, 8%) while the highest proportion of DS was provided for inhalation medication (n=78,541/126,588, 62%), oral drops (52,085/98,221, 53%) and tablets, capsules, suppositories (2,828,617/6,176,585, 45.8%), in their respective drug form group (Table 2).

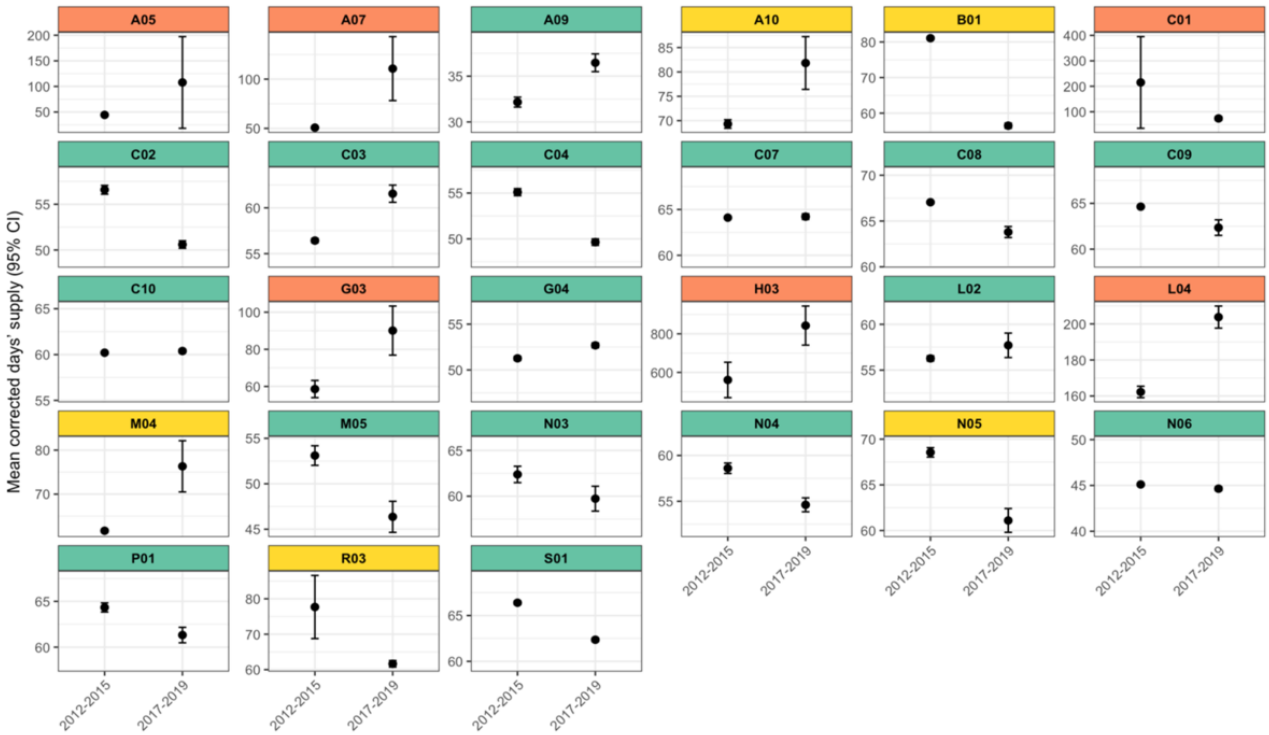
For tablets, capsules, and suppositories, 38.9% (2,402,392/6,176,585) of prescriptions lacked daily dose information and were therefore imputed using the mode-based imputation described earlier. The mode table consisted of 1002 active substances and dosage amount combinations, out of which 60.1% (602/1002) of combinations had the value of once per day, and 24.7% (245/1002) of these combinations had the value of twice per day. In cases where both a calculated DS and provided DS information were available, the smaller value was used out of the 2. As a result, 15.1% (933,710/6,176,585) of prescriptions in single-dose medications were assigned DS provided by the doctor. For other drug forms, when the DS could not be calculated, instead of using the daily dose, we imputed the DS based on the specific drug form. The proportion of imputed prescriptions ranged from 13.2% (12,961/98,221) for oral drops to 92% (19,453/21,145) for intravaginal contraceptives and 96.9% (80,266/82,867) for insulin injections (Table 2; Multimedia Appendix 2). As a result of applying different imputation approaches, we successfully found the DS for 98.3% (7,415,347/7,544,892) of dispensed prescriptions. For 1.7% (129,545/7,544,892) of prescriptions, DS could neither be imputed nor calculated.

To evaluate the impact of data cleaning and imputation, the mean DS values and medication adherence of 147 active substances belonging to 27 ATC therapeutic subgroups were calculated.

Comparison of the mean DS values between the period where most DS were imputed (2012-2015) and the period with more complete daily dose and DS data (2017-2019) revealed that the observed differences in mean DS were generally small and, in most cases, did not exceed the difference of 7 days (Figure 2). This indicates a high level of consistency between imputed and calculated values. Nonetheless, some ATC groups with larger deviations were present, for example, thyroid therapy (H03) and cardiac therapy (C01).

XSL•FO
RenderX

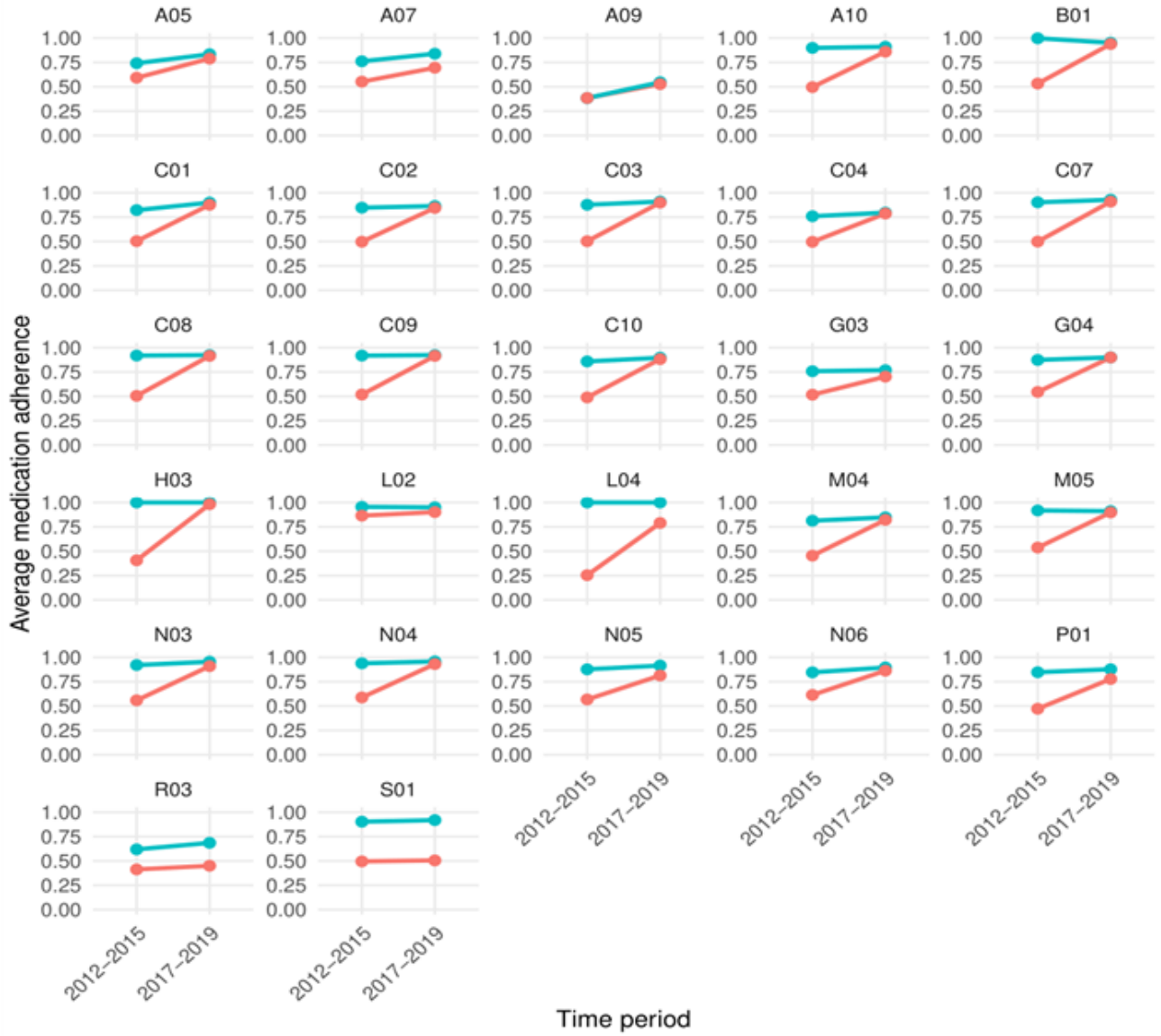
Figure 2. The mean value of DS and 95% CIs for a period where DS was mostly imputed (years 2012-2015) and mostly calculated based on data provided by doctors (years 2017-2019) by Anatomical Therapeutic Chemical subgroups. Green: ≤7 mean difference; yellow: 7-30 mean; red: > 30 mean difference.A05: bile and liver therapy; A07: antidiarrheals, intestinal anti-inflammatory and anti-infective agents; A09: digestives, including enzymes; A10: drugs used in diabetes; B01: antithrombotic agents; C01: cardiac therapy; C02: antihypertensives; C03: diuretics; C04: peripheral vasodilators; C07: beta blocking agents; C08: calcium channel blockers; C09: agents acting on the renin-angiotensin system; C10: lipid modifying agents; G03: sex hormones and modulators of the genital system; G04: urologicals; H03: thyroid therapy; L02: endocrine therapy; L04: immunosuppressants; M04: antigout preparations; M05: drugs for treatment of bone diseases; N03: antiepileptics; N04: anti-Parkinson drugs; N05: psycholeptics; N06: psychoanaleptics; P0: antiprotozoals, R03: drugs for obstructive airway diseases; S01: ophthalmologicals.



The medication adherence comparison shows that for most of the drug groups, the difference between 2 observed time periods was more distinct in the baseline dataset compared with the corrected dataset, indicating that the applied correction methods had lessened the stark contrast (Figure 3; Multimedia Appendix 1). For the multiple-dose medications, such as drugs for obstructive airway diseases (R03) and ophthalmologicals (S01), there was no distinction between the observed time periods. However, the medication adherence improved similarly for both time periods. At the same time, some adherence measures, such as the thyroid therapy (H03), showed 100% of medication adherence.

XSL•FO
RenderX

Figure 3. The average medication adherence by time period and Anatomical Therapeutic Chemical subgroups. Blue: corrected dataset; red: baseline dataset. A05: bile and liver therapy; A07: antidiarrheals, intestinal anti-inflammatory and anti-infective agents; A09: digestives, including enzymes; A10: drugs used in diabetes; B01: antithrombotic agents; C01: cardiac therapy; C02: antihypertensives; C03: diuretics; C04: peripheral vasodilators; C07: beta blocking agents; C08: calcium channel blockers; C09: agents acting on the renin-angiotensin system; C10: lipid modifying agents; G03: sex hormones and modulators of the genital system; G04: urologicals; H03: thyroid therapy; L02: endocrine therapy; L04: immunosuppressants, M04: antigout preparations; M05: drugs for treatment of bone diseases; N03: antiepileptics; N04: anti-Parkinson drugs; N05: psycholeptics; N06: psychoanaleptics; P01: antiprotozoals; R03: drugs for obstructive airway diseases; S01: ophthalmologicals.



Discussion

Principal Results

Accurate estimation of DS is essential for assessing medication adherence and conducting pharmacoepidemiological research. This paper set out to improve the completeness and precision of prescription data. We developed and implemented a multistep data cleaning and imputation approach to address missing or incomplete information in prescription records, specifically targeting the derivation of DS. Unlike previous studies, which typically focused on a single drug class or drug form [4,6-9,11,13-18,20,26,27], our work aimed to determine DS values for all prescriptions. By applying a combination of rule-based calculations, statistical imputation, and domain knowledge, we were able to assign DS values to almost all of the 7.5 million dispensed prescriptions included in this study dataset. This demonstrates the feasibility of using hybrid methods in large-scale, real-world prescription datasets, particularly when working with data mapped to standardized models.

The mean value comparison indicates a high level of consistency between imputed and observed values in most medication groups. These findings suggest that the imputation strategy provides a reasonable approximation of real-world prescription lengths and preserves the overall patterns in medication use. Nonetheless, isolated ATC groups with larger deviations show that imputed values may still under- or overestimate treatment duration in specific cases. Furthermore, it is evident that the imputation process did not systematically shift the DS estimates upward or downward. This suggests that the observed

XSL•FO
RenderX

differences in DS between time periods might also be driven by external factors, such as changes in available package sizes or evolving prescribing practices.

The medication adherence calculated on the baseline database and corrected database suggests that our approach significantly improved the adherence estimates. For most drug groups, it became evident that once imputation methods were applied, the adherence estimates for the 2012-2015 period aligned more closely with those from 2017-2019, when daily dose reporting became mandatory for physicians and the overall database quality improved. At the same time, the adherence values for the 2017-2019 period for both datasets remained similar for most drug classes. This supports the hypothesis that earlier data underrepresented medication availability due to incomplete documentation, and that our imputation procedures improved temporal consistency and data reliability.

Some kind of imputation was needed in each observed drug class. The predominance of tablets, capsules and suppositories in the dataset enabled the use of daily dose–based imputations for a substantial portion of the records. Imputing 1 dose per day has proven to be an effective way to address the missing data for some oral drugs in a previous study with heart medications [6]. Although this approach was also applicable in our dataset, the mode DS values revealed that such a uniform approach was unsuitable for 39.9% (400/1002) of single dose medications imputation values. Therefore, using the mode-table imputation method helped to identify the most common daily doses per active substance and use this information in DS calculations. Although our approach significantly improves the data quality, it is not flawless. For example, the medication adherence calculated for thyroid therapy (H03) on the corrected database was 100%. Thus, raising the suspicion that it might be overestimated and the imputed daily doses were too small. It could be hypothesized that for some medications more than others, the individual treatment regimens may differ, and therefore, it is difficult to identify the most common dose to impute. When the mode value used in imputation is lower than the next most popular value, then it can result in a higher DS value and, therefore, better medication adherence values.

For non–single-dose drug forms, such as creams, drops, and gels, imputation based on SPCs and provided DS proved useful, though these methods are inherently less precise due to variability in usage patterns and dosing recommendations. In 2 drug classes, obstructive airway diseases (R03) and ophthalmologicals (S01), the medication improved for both time periods. One reason for this could be that in the baseline dataset, a rough imputation of 30 days was applied to all missing daily doses, while in this study, the imputed DS was more active substance specific and based on national SPCs.

In addition to imputing missing data, there is sometimes a question of the plausibility of provided DS that are given by a doctor with the prescription [5,19]. The question arises whether and how we assess the DS prescribed by physicians, and whether this should be compared with the calculated duration. Until 2016, our prescription system allowed manual entry of the duration, and errors can often occur. The most common and noticeable errors occurred when repeat prescriptions were issued, and each was assigned a DS of 180 days, which in fact represents the combined length of 3 prescriptions. This stems from the common practice of doctors providing 3 refill prescriptions per medication. These kinds of problems were easy to notice and fix, but more complex errors were harder to detect. Another example concerns vaccines and implants, for which no standard exists for indicating the DS on the prescription, and each doctor records it according to their own discretion. As a result of incorrect entries, some prescriptions may end up with an inaccurate DS value. Despite these challenges, our imputation methods combined with domain knowledge ensured reasonable estimates of DS across diverse drug forms.

Our study also underscores the importance of a user design and information architecture of the prescription database. It clearly emerged from this study that before the summer of 2016, when there was no requirement to record dosing instructions for medications, the data quality was lower and different methods were needed to backfill this information retrospectively. Therefore, to collect accurate data, more effort should be paid to the architecture of the system to ensure that all necessary data will be inserted and stored as correctly as possible. Moreover, there is a need to raise awareness among doctors on the importance of data quality and its effect on evaluating health care services and medication adherence. For example, we identified in the dataset that sometimes there is an inconsistency in the units of prescribed medications within prescriptions issued by doctors—one might use units (eg, 1 tablet) on one prescription and the amount of active substance (eg, 10 mg) on another. Moreover, some doctors have a practice to renew the old prescription without changing the dosing information, even when the dosing regimen changes. This all impacts the data quality, as detecting such cases from the data is very difficult, if not impossible. Therefore, beyond seeking sophisticated imputation methods to address the missing data, we should also consider improving the prescription systems and informing and educating the doctors who enter this data. More complete and accurate records would provide a better foundation for secondary use of prescription data in the future.

Limitations

One limitation of our study is that there was no golden standard or reference database to compare the results with. Although the amount of missing data substantially reduced in the period 2017-2019 due to the changes in the prescription system, some inaccuracies due to the human component remained. However, it could be argued that the baseline data from 2017-2019 gives a considerably good indication of actual prescription patterns.

It is also important to acknowledge that imputation techniques possess inherent limitations and may not invariably produce fully accurate estimates. To construct a mode table of daily doses, a certain proportion of presumably correct prescription data is required; otherwise, it cannot be compiled. Furthermore, in some drug classes where the dosing recommendations vary based on the severity of disease, the mode table did not seem to be the best approach, as the medication adherence calculated on the corrected database was unrealistically high. Potentially, some ML methods could be more effective in such cases where

XSL•FO
RenderX

mode tables fail, but this warrants further investigation. Moreover, ML methods should also be applied to other injections and less prevalent drug forms, which in this study were excluded.

Conclusions

In summary, our study demonstrated that with a carefully designed imputation pipeline where data-driven imputation is combined with domain knowledge and literature information, it is possible to meaningfully improve the quality of prescription datasets and generate more accurate and consistent adherence metrics across various drug forms. Nonetheless, future efforts should continue to refine imputation techniques, incorporate ML approaches where appropriate, and expand validation efforts using external benchmarks or clinical outcomes.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to legal restrictions on sharing deidentified data. According to legislative regulation and data protection law in Estonia, the authors cannot publicly release the data received from the health data registries in Estonia. However, the data can be requested by completing necessary applications in order to carry out research or an evaluation of public interest and acquiring the permission of the controller of the databases.

Funding

This study was co-funded by the European Union and the Estonian Ministry of Education and Research via project TEM-TA72 and Estonian Research Council (grant PRG1844).
This project has received funding from the European Union's Horizon Europe research and innovation program (grant agreement No 101060011). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This research was co-funded by the European Union through the European Regional Development Fund (Project No. 2021-2027.1.01.24-0444). This research was funded by the Estonian Ministry of Education and Research (Teaming for Excellence).

Authors' Contributions

Conceptualization and methodology: MM, KM, RK, MO, JH, and NU
Data curation: MM, MO, and ST
Formal analysis: MM and KM
Funding and resources: RK, SR, and JV
Investigation: MM, KM, and HK
Software: MM, JH, and MO
Supervision: KM and RK
Visualization: MM, KM, and RK
Writing – review and editing: MM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Graphs of yearly medication adherence (2012-2019) showing baseline dataset and corrected dataset. Stratified by ATC code.
[PDF File (Adobe PDF File), 1742 KB - ojphi_v18i1e83465_app1.pdf]

Multimedia Appendix 2
Prescription distribution by drug groups and days' supply establishing methods, for the full dataset (2012-2019).
[PDF File (Adobe PDF File), 95 KB - ojphi_v18i1e83465_app2.pdf]

References

1. Miao Z, Sealey MD, Sathyanarayanan S, Delen D, Zhu L, Shepherd S. A data preparation framework for cleaning electronic health records and assessing cleaning outcomes for secondary analysis. Information Systems 2023 Jan;111:102130. [doi: 10.1016/j.is.2022.102130]
2. Vollmer WM, Xu M, Feldstein A, Smith D, Waterbury A, Rand C. Comparison of pharmacy-based measures of medication adherence. BMC Health Serv Res 2012 Jun 12;12:155 [FREE Full text] [doi: 10.1186/1472-6963-12-155] [Medline: 22691240]
3. Wettermark B, Zoëga H, Furu K, Korhonen M, Hallas J, Nørgaard M, et al. The Nordic prescription databases as a resource for pharmacoepidemiological research--a literature review. Pharmacoepidemiol Drug Saf 2013 Jul;22(7):691-699. [doi: 10.1002/pds.3457] [Medline: 23703712]

XSL•FO
RenderX

4. Lum KJ, Newcomb CW, Roy JA, Carbonari DM, Saine ME, Cardillo S, et al. Evaluation of methods to estimate missing days' supply within pharmacy data of the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). Eur J Clin Pharmacol 2017 Jan;73(1):115-123. [doi: 10.1007/s00228-016-2148-4] [Medline: 27787616]
5. Xu C, Ferrante SA, Fitzgerald T, Pericone CD, Wu B. Inconsistencies in the days supply values reported in pharmacy claims databases for biologics with long maintenance intervals. J Manag Care Spec Pharm 2023 Jan;29(1):90-100 [FREE Full text] [doi: 10.18553/jmcp.2023.29.1.90] [Medline: 36580125]
6. Ung D, Dalli LL, Lopez D, Sanfilippo FM, Kim J, Andrew NE, et al. Assuming one dose per day yields a similar estimate of medication adherence in patients with stroke: An exploratory analysis using linked registry data. Br J Clin Pharmacol 2021 Mar;87(3):1089-1097 [FREE Full text] [doi: 10.1111/bcp.14468] [Medline: 32643250]
7. Dalli LL, Kilkenny MF, Arnet I, Sanfilippo FM, Cummings DM, Kapral MK, et al. Towards better reporting of the proportion of days covered method in cardiovascular medication adherence: A scoping review and new tool TEN-SPIDERS. Br J Clin Pharmacol 2022 Oct;88(10):4427-4442 [FREE Full text] [doi: 10.1111/bcp.15391] [Medline: 35524398]
8. Saukkosalmi P, Kankaanranta H, Vähätalo I, Sillanmäki L, Sumanen M. Defined daily dose definition in medication adherence assessment in asthma. Eur Clin Respir J 2023;10(1):2207335 [FREE Full text] [doi: 10.1080/20018525.2023.2207335] [Medline: 37139181]
9. Nielsen S, Gisev N, Bruno R, Hall W, Cohen M, Larance B, et al. Defined daily doses (DDD) do not accurately reflect opioid doses used in contemporary chronic pain treatment. Pharmacoepidemiol Drug Saf 2017 May;26(5):587-591. [doi: 10.1002/pds.4168] [Medline: 28101968]
10. ATC/DDD toolkit. World Health Organization. URL: https://www.who.int/tools/atc-ddd-toolkit [accessed 2025-07-01]
11. Tapply I, Broadway DC. Improving Adherence to Topical Medication in Patients with Glaucoma. Patient Prefer Adherence 2021;15:1477-1489 [FREE Full text] [doi: 10.2147/PPA.S264926] [Medline: 34239297]
12. Sinnott SJ, Polinski JM, Byrne S, Gagne JJ. Measuring drug exposure: concordance between defined daily dose and days' supply depended on drug class. J Clin Epidemiol 2016 Jan;69:107-113 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.05.026] [Medline: 26146090]
13. Friedman DS, Quigley HA, Gelb L, Tan J, Margolis J, Shah SN, et al. Using Pharmacy Claims Data to Study Adherence to Glaucoma Medications: Methodology and Findings of the Glaucoma Adherence and Persistency Study (GAPS). Invest. Ophthalmol. Vis. Sci 2007 Nov 01;48(11):5052-5057. [doi: 10.1167/iovs.07-0290] [Medline: 17962457]
14. Assem AS, Fekadu SA, Yigzaw AA, Nigussie ZM, Achamyeleh AA. Level of Glaucoma Drug Adherence and Its Associated Factors Among Adult Glaucoma Patients Attending Felege Hiwot Specialized Hospital, Bahir Dar City, Northwest Ethiopia. Clin Optom (Auckl) 2020;12:189-197 [FREE Full text] [doi: 10.2147/OPTO.S274850] [Medline: 33149714]
15. Muñoz-Villegas P, Martínez-Bautista H, Olvera-Montaño O. Determinants of adherence to treatment in patients with ophthalmic conditions. Expert Rev Clin Pharmacol 2023;16(12):1249-1259 [FREE Full text] [doi: 10.1080/17512433.2023.2279740] [Medline: 37978952]
16. Gupta G, Mallefet P, Kress DW, Sergeant A. Adherence to topical dermatological therapy: lessons from oral drug treatment. Br J Dermatol 2009 Aug;161(2):221-227. [doi: 10.1111/j.1365-2133.2009.09253.x] [Medline: 19548959]
17. Fenske C, Boytsov N, Guo J, Dawson Z. Prescription Market Share and Treatment Patterns in Atopic Dermatitis: A Retrospective Observational Study Using US Insurance Claims. Adv Ther 2022 May;39(5):2052-2064 [FREE Full text] [doi: 10.1007/s12325-022-02071-y] [Medline: 35287230]
18. Lo A, Lovell KK, Greenzaid JD, Oscherwitz ME, Feldman SR. Adherence to treatment in dermatology: Literature review. JEADV Clinical Practice 2024 Feb 14;3(2):401-418. [doi: 10.1002/jvc2.379]
19. Singh H, Mani S, Espadas D, Petersen N, Franklin V, Petersen LA. Prescription errors and outcomes related to inconsistent information transmitted through computerized order entry: a prospective study. Arch Intern Med 2009 May 25;169(10):982-989 [FREE Full text] [doi: 10.1001/archinternmed.2009.102] [Medline: 19468092]
20. Burden AM, Huang A, Tadrous M, Cadarette SM. Variation in the days supply field for osteoporosis medications in Ontario. Arch Osteoporos 2013;8:128. [doi: 10.1007/s11657-013-0128-1] [Medline: 23475734]
21. Oja M, Tamm S, Mooses K, Pajusalu M, Talvik H, Ott A, et al. Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned. JAMIA Open 2023 Dec;6(4):ooad100 [FREE Full text] [doi: 10.1093/jamiaopen/ooad100] [Medline: 38058679]
22. The Register of Medicinal Products. The Register of Medicinal Products. URL: https://www.ravimiregister.ee/ [accessed 2025-07-01]
23. Estonian legislation. Riigi Teataja. URL: https://www.riigiteataja.ee/akt/130042025002?leiaKehtiv [accessed 2025-08-01]
24. Dima AL, Dediu D. Computation of adherence to medication and visualization of medication histories in R with AdhereR: Towards transparent and reproducible use of electronic healthcare data. PLoS One 2017;12(4):e0174426 [FREE Full text] [doi: 10.1371/journal.pone.0174426] [Medline: 28445530]
25. Holm J. AdherenceFromOMOP. GitLab. URL: https://gitlab.cs.ut.ee/johanneh/AdherenceFromOMOP [accessed 2025-05-01]
26. Nelson AJ, Pagidipati NJ, Bosworth HB. Improving medication adherence in cardiovascular disease. Nat Rev Cardiol 2024 Jun;21(6):417-429. [doi: 10.1038/s41569-023-00972-1] [Medline: 38172243]

XSL•FO
RenderX

27. Tibble H, Sheikh A, Tsanas A. Estimating medication adherence from Electronic Health Records: comparing methods for mining and processing asthma treatment prescriptions. BMC Med Res Methodol 2023 Jul 12;23(1):167 [FREE Full text] [doi: 10.1186/s12874-023-01935-3] [Medline: 37438684]

Abbreviations

ATC: Anatomical Therapeutic Chemical
DDD: defined daily dose
DS: days' supply
CMA: continuous multiple interval measures of medication availability
ML: machine learning
SPC: Summary of Product Characteristics

Edited by L Zhou; submitted 03.Sep.2025; peer-reviewed by MK Ghanta, I Aznar-Lou; comments to author 21.Nov.2025; revised version received 12.Dec.2025; accepted 29.Dec.2025; published 11.Feb.2026.

Please cite as:
Malk M, Mooses K, Oja M, Holm J, Keidong H, Umov N, Tamm S, Reisberg S, Vilo J, Kolde R
A Comprehensive Approach to Days' Supply Estimation in a Real-World Prescription Database: Algorithm Development and Validation Study
Online J Public Health Inform 2026;18:e83465
URL: https://ojphi.jmir.org/2026/1/e83465
doi:10.2196/83465
PMID:

©Maria Malk, Kerli Mooses, Marek Oja, Johannes Holm, Hanna Keidong, Nikita Umov, Sirli Tamm, Sulev Reisberg, Jaak Vilo, Raivo Kolde. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 11.Feb.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Knowledge, Attitudes, and Perceptions Regarding Mini Implants Among Orthodontic Practitioners in the Kurdistan Region: Cross-Sectional Study

Soran M Hameed[1*], BDS; Aras M Rauf[2*], PD

[1]Orthodontic Training Center, Sulaymaniyah, Kurdistan Region, Iraq

[2]Department of Dentistry, College of Dentistry, University of Sulaimani, Sulaminyay, Kurdistan Region, Iraq

[*]all authors contributed equally

Corresponding Author:
Soran M Hameed, BDS
Orthodontic Training Center
Goran Street
Sulaymaniyah, Kurdistan Region, 46001
Iraq
Phone: 964 7701415353
Email: sorandentist@gmail.com

Abstract

Background: Mini implants, or temporary anchorage devices, have transformed modern orthodontic practice by offering stable, minimally invasive anchorage for complex tooth movements. Despite their proven effectiveness, their use varies widely across regions, often influenced by clinicians' knowledge, confidence, and training.

Objective: This study explored the knowledge, attitudes, and perceptions of orthodontic professionals regarding mini implant use in the Kurdistan Region, Iraq, and examined how experience and professional background shape their adoption in daily clinical work.

Methods: A cross-sectional online survey was conducted between April 2025 and September 2025 among orthodontic faculty and postgraduate trainees from 7 dental colleges in the Kurdistan Region. The validated questionnaire assessed participants' demographic details and 3 key domains—knowledge, perceptions, and attitudes—using Likert-scale responses. Data were analyzed using SPSS (version 28.0) and the Mann-Whitney *U*, Kruskal-Wallis, and Spearman correlation tests, with significance set at $P \leq .05$.

Results: A total of 175 orthodontic practitioners completed the survey (n=100, 57.1% postgraduate trainees; n=75, 42.9% faculty members). Postgraduate trainees demonstrated significantly higher knowledge (mean 3.66, SD 0.49 vs mean 3.16, SD 0.48; $P=.01$) and perception (mean 3.29, SD 0.60 vs mean 2.39, SD 0.58; $P=.02$) scores. Immediate loading was preferred by 80% (80/100) of the postgraduate trainees compared with 40% (30/75) of faculty members ($P=.001$), whereas radiographic guidance was selected as the safest placement method by 75% (75/100) of postgraduate trainees vs 40% (30/75) of faculty members ($P<.001$). Younger clinicians (aged <35 years) and those with less than 5 years of experience showed significantly higher perception scores ($P=.01$). Knowledge, perception, and attitude were strongly correlated ($r=0.74$; $P<.001$), indicating that increased understanding promotes more positive attitudes toward mini screw use.

Conclusions: Orthodontists in the Kurdistan Region generally hold favorable views toward mini implants, yet differences in confidence and hands-on experience remain evident across generations.

(Online J Public Health Inform 2026;18:e86896) doi:10.2196/86896

KEYWORDS

orthodontic mini implants; temporary anchorage devices; orthodontists; knowledge; attitude; perception; skeletal anchorage; Kurdistan Region; cross-sectional study

XSL•FO
RenderX

Introduction

Orthodontic implants trace their origins to the mid-20th century, when Gainsforth and Higley [1] first described the use of screws as auxiliary anchorage devices for tooth movement. Although their initial attempts achieved limited implant survival, these pioneering experiments laid the foundation for future advancements. Over the subsequent decades, orthodontic research focused intensively on developing reliable intraoral anchorage systems, leading to significant innovations in implant design and application for both prosthodontic and orthognathic purposes [2].

The evolution of mini screw technology accelerated in the 1990s, when Kanomi [3] introduced screws specifically designed for orthodontic use. Improvements in design geometry, size, and alloy composition enhanced both biocompatibility and mechanical stability, making mini screws more accessible and predictable for routine clinical practice. Several specialized systems have emerged, including the Aarhus mini implant, Spider Screw, and AbsoAnchor microimplant, reflecting global efforts to optimize skeletal anchorage [4].

Achieving optimal anchorage control remains one of the most persistent challenges in orthodontics, directly influencing treatment predictability and success. The introduction of skeletal anchorage systems, particularly orthodontic mini screws, has represented a paradigm shift in anchorage reinforcement, offering a stable, compliance-free alternative for complex tooth movements. Mini implants are appreciated for their versatility, minimal invasiveness, ease of placement and removal, and capacity for immediate or early loading. These advantages have broadened their indications across numerous nonsurgical orthodontic applications such as en masse retraction, molar distalization, open bite correction, maxillary expansion, and the management of complex malocclusions. Clinical success is primarily linked to mini screw stability, which depends on bone quality, insertion torque and depth, and biomaterial characteristics. Reported survival rates often exceed 80% to 90%, affirming their reliability as adjuncts in contemporary orthodontic treatment [2].

Despite these benefits, the global adoption of mini implants remains uneven, influenced by practitioners' knowledge, clinical experience, and access to specialized training. Reported barriers include screw loosening (up to 83%), soft tissue irritation (59.7%), procedural failure, and concerns about patient comfort and oral hygiene. These factors underscore the ongoing need for technical proficiency, standardized guidelines, and practitioner confidence [5,6].

Recent studies show that orthodontists' willingness to use mini screws is shaped by variables such as professional experience, continuing education, gender, and practice setting. A study revealed that over 90% of orthodontists who use mini screws report satisfactory or highly satisfactory treatment outcomes, yet concerns about cost, procedural complexity, and potential complications persist. Clinical factors rather than patient demographics such as age or gender are the primary determinants of mini screw success [7,8].

Worldwide, the use of mini screws continues to expand, supported by strong clinical evidence and favorable outcomes. However, a gap remains between early-career and senior orthodontists regarding optimal placement sites, complication management, and biomechanical integration. Addressing these disparities requires structured education, mentorship, and evidence-based training to support effective diffusion of this technology [9,10].

In response to these trends, this study investigated the current knowledge, attitudes, and perceptions (KAP) related to orthodontic mini screw use among orthodontists in the Kurdistan Region, Iraq. This study also examines the demographic and professional factors influencing clinicians' readiness to incorporate mini screws into routine practice. It is hypothesized that there will be significant variation in KAP scores based on experience level, training background, practice environment, and access to professional development opportunities. Specifically, this research aimed to (1) assess orthodontists' knowledge of mini screw indications, benefits, and complications; (2) evaluate their actual clinical practices, including preferred insertion sites, perceived success rates, and common barriers; (3) explore attitudes toward mini screws compared with conventional anchorage methods; and (4) identify the major educational and technical gaps limiting their optimal use.

This investigation represents the first comprehensive regional evaluation of orthodontists' readiness to adopt mini implant technology, with findings expected to inform targeted continuing education programs, enhance clinical quality, and advance evidence-based orthodontic care across the Kurdistan Region.

Methods

Study Design and Setting

This cross-sectional, descriptive, questionnaire-based study was conducted from April 1, 2025, to September 1, 2025, by the Medical Research Scientific and Ethical Committee in collaboration with orthodontic faculty and postgraduate practitioners from dental colleges across the Kurdistan Region of Iraq. Eligible participants included registered orthodontists in the Kurdistan Dental Association and postgraduate trainees actively studying in accredited dental institutions within the region.

Ethical Considerations

The study protocol was approved by the Kurdistan Higher Council of Medical Specialties Medical Research Scientific and Ethical Committee. It was conducted in accordance with the ethical principles of the 2013 Declaration of Helsinki. All participation was entirely voluntary, and the study objectives and confidentiality safeguards were clearly stated before participation. Informed consent was explicitly obtained before the start of the survey. Participants did not receive any compensation for taking part in the study. To protect participants' anonymity, the survey did not collect any personally identifying information, and all responses were accessible solely to the principal investigator.

XSL•FO
RenderX

The Questionnaire

Overview

The questionnaire was constructed following the CHERRIES (Checklist for Reporting Results of Internet E-Surveys; Typeform) guidelines, ensuring concise, validated, and ethically sound data collection. The first section of the questionnaire included a demographic statement detailing the study objectives, eligibility criteria, voluntariness of participation, and contact details of the principal investigator.

The questionnaire was systematically structured into 3 consecutive sections (A, B, and C) to ensure that, upon completion of the study, a comprehensive assessment and understanding could be achieved regarding the respondents' characteristics and their knowledge about, perceptions of, and clinical attitudes toward orthodontic mini screws.

Demographic Section

This section collected information on participants' age, gender, academic designation, region of practice, years of experience, and involvement in clinical orthodontic work.

Knowledge Assessment

This subsection included multiple-choice and Likert-scale items that evaluated participants' understanding of mini implant indications, benefits, complications, loading protocols, and stability principles.

Perception Assessment

Perception was assessed using Likert-scale items measuring participants' views on clinical safety, barriers to use, success rates, loading preferences, and confidence regarding mini screw application.

Attitude Assessment

Attitude items explored respondents' willingness to use mini implants, openness to adopting new techniques, preferred clinical approaches, and readiness for training and skill development.

Pretesting and Content Validity of the Questionnaire

The questionnaire was developed through the collaborative input of a panel consisting of 2 orthodontists, 1 academic statistician, and 1 researcher with over 15 years of research experience, all independent from the study's participant pool. Content validity was established via expert review, with the Aiken *V* statistic determined to be 0.89, indicating a high level of agreement among reviewers. To further refine the instrument, a pilot test was conducted with 25 orthodontists and postgraduate trainees not included in the final sample.

The internal consistency of the questionnaire was assessed using the Cronbach α, yielding coefficients of 0.85 for the knowledge domain and 0.91 for attitude and perception items. Test-retest reliability was evaluated by readministering the instrument to the pilot group after a 3-week interval, resulting in an interobserver κ coefficient of 0.88. These procedures ensured strong measurement validity and reliability, with the finalized version distributed to a sample of approximately 120 orthodontists and postgraduate participants, aligning with standard methodological frameworks for cross-sectional KAP surveys in dental research.

Sample Size Estimation

The sample size was estimated using the online tool provided by the Calculator website [11], ensuring an evidence-based approach for determining the appropriate participant number [12]. According to records from the Kurdistan Dental Association, the total number of orthodontists and postgraduate trainees practicing in the region was approximately 210 at the time of the study. Using this population figure, a statistical power analysis was conducted to calculate the minimum required sample size.

Assuming an anticipated response proportion of 50% on mini screw–related knowledge and attitudes with a 5% margin of error and a confidence level set at 95%, the recommended minimum sample size was 132. Considering an estimated response rate of 75% based on prior regional survey experiences, the questionnaire was distributed to 175 practitioners to achieve a representative and statistically robust sample.

Study Participants and Data Collection Procedure

The roster of dental institutions and relevant contact details were obtained from the official records of the Kurdistan Dental Association and the Ministry of Higher Education and Scientific Research of the Kurdistan Region. Formal correspondence, endorsed by the head of the orthodontics department, was sent to dental colleges and teaching hospitals across the northern provinces requesting updated lists of both orthodontic faculty and postgraduate trainees, including email and telephone contact information. Among the 7 dental colleges in the Kurdistan Region (with representation from Erbil, Sulaymaniyah, Duhok, Halabja, and the surrounding areas), a total of 123 postgraduate orthodontic trainees and 87 faculty members were identified, for a combined total of 210 eligible participants—a postgraduate trainee–to–faculty member ratio of approximately 1.4:1.

A convenience nonprobabilistic sampling approach was used. Google Form invitations, including instructions and electronic consent, were distributed proportionally to 175 individuals (n=100, 57.1% postgraduate students and n=75, 42.9% faculty members) via email and professional WhatsApp groups, reflecting the regional distribution and the desired sample size as determined via power analysis. Participants had the opportunity to complete the questionnaire once. Duplicate and incomplete responses were screened out, and reminder messages were sent 2 weeks after the initial invitation to optimize the response rate. Inclusion was limited to orthodontic faculty and postgraduate trainees, whereas undergraduate students, those who did not provide consent, and incomplete submissions were excluded from the study.

Scoring Criteria

Overview

In the analysis, each Likert-scale question except for designated demographic and multiple-choice questions was scored as follows: "agree" was assigned a value of 1, "neutral" was assigned a value of 0, and "disagree" was assigned a value of −1. For the knowledge section, comprising 8 Likert-scale items,

XSL•FO
RenderX

the maximum achievable score was 8. Similarly, the perception and attitude sections each contained 6 Likert-scale items, yielding maximum possible scores of 6 for each domain. Summary metrics, including means and SDs, were calculated for the knowledge, perception, and attitude scores to enable quantitative comparison among respondents.

Knowledge Scoring

Knowledge items were scored using a 3-point Likert scale where “agree”=1, “neutral”=0, and “disagree”=−1.

The knowledge section contained 8 items, yielding a possible score range of −8 to +8, with higher scores indicating better knowledge regarding mini screw use.

Perception Scoring

Perception items were also scored using the same Likert scale (“agree”=1, “neutral”=0, and “disagree”=−1).

This section included 6 items, allowing for a maximum perception score of +6.

Higher values reflected a more positive perception of mini screw application.

Attitude Scoring

Attitude items followed the same scoring pattern (1, 0, and −1). Six attitude items were included, with a maximum total possible score of +6. Higher scores indicated a more favorable attitude toward incorporating mini screws into orthodontic practice.

Statistical Analysis

Data from the online questionnaire platform were checked and extracted. Any apparent disparity or error was spotted, discussed, and reconciled to ensure that incomplete or inconsistent entries were successfully eliminated from the final analysis. The SPSS software (version 28.0; IBM Corp) was used for the statistical analyses.

The normality of continuous variables was checked via the Shapiro-Wilk test. Descriptive statistics such as frequencies and percentages were used to summarize sociodemographic characteristics and the distribution of responses. Group comparisons applied nonparametric tests: the Mann-Whitney *U* test for dichotomous variables (such as gender) and the Kruskal-Wallis test for multigroup comparisons by experience level, academic position, or region. Where significant differences emerged, Dunn post hoc tests were used in pairwise analyses. KAP scores were also analyzed using these nonparametric methods. Associations between ordinal variables such as KAP scores and years of experience were tested using the Spearman rank-order correlation coefficient (Spearman ρ). The threshold for statistical significance was set at $P \leq .05$ across all analyses.

Results

Participant Demographics

Table 1 shows the baseline demographic characteristics of the study respondents. Of the total sample size of 175 valid responses, 100 (57.1%) were from postgraduate students, whereas 75 (42.9%) were faculty members. Participants were recruited from different provinces, and their distribution across Kurdistan dental colleges varied by region. The respondent population was balanced between men and women, and most were in the age group of 26 to 35 years. A total of 66.9% (117/175) indicated less than 5 years of professional experience, and 12.6% (22/175) indicated more than 10 years of professional experience. The study sample represented orthodontic practitioners from several provinces in the Kurdistan Region, with participants from Erbil, Sulaymaniyah, and Duhok, reflecting a diverse subregional distribution. A total of 87% (65/75) of the faculty members were in private practice.

XSL•FO
RenderX

Table 1. Baseline characteristics of the study participants (n=175).

Demographic variable and category	Participants, n (%)
Age group (y)	
<25	29 (16.6)
26-35	108 (61.7)
36-45	29 (16.6)
>45	9 (5.1)
Sex	
Male	88 (50.3)
Female	87 (49.7)
Academic designation	
Postgraduate students	100 (57.1)
Faculty members	75 (42.9)
Experience (y)	
<5	117 (66.9)
5-10	36 (20.6)
>10	22 (12.6)
Private practice (faculty only; n=75)	
Yes	65 (86.7)
No	10 (13.3)

Knowledge Results

Significant differences in knowledge-related items were observed between the 2 groups. For primary stability mechanisms, 60% (60/100) of postgraduate students attributed mini screw stability to bone quality, whereas 55% (41/75) of faculty members emphasized cortical engagement (*P*=.002).

Regarding success rates, 70% (70/100) of postgraduate students reported 80% to 90% success rates, whereas 65% (49/75) of faculty members reported success rates above 90% (*P*=.03).

For clinical indications, anterior retraction was selected by 55% (55/100) of postgraduate students compared with 27% (20/75) of faculty members, whereas molar intrusion or extrusion was selected by 45% (34/75) of faculty members (*P*=.02). With respect to placement safety, 75% (75/100) of postgraduate students preferred radiographic imaging, whereas 55% (41/75) of faculty members favored safe angulation techniques (*P*<.001). A detailed comparison of knowledge-related responses is shown in Table 2.

XSL•FO
RenderX

Table 2. Comparison of responses between postgraduate students and faculty members using the chi-square test.

Variable and response	Postgraduate students (n=100), n (%)	Faculty members (n=75), n (%)	Chi-square (*df*)	*P* value
Primary stability mechanism			9.4 (2)	.002
Bone quality and density	60 (60)	26 (34.7)		
Cortical engagement	30 (30)	41 (54.7)		
Mechanical interlocking	10 (10)	8 (10.7)		
Success rate (%)			7.1 (2)	.03
80-90	70 (70)	22 (29.3)		
>90	20 (20)	49 (65.3)		
<80	10 (10)	4 (5.3)		
Clinical indication			5.9 (2)	.02
Anterior retraction	55 (55)	20 (26.7)		
Molar intrusion or extrusion	25 (25)	34 (45.3)		
Others	20 (20)	21 (28)		
Placement safety			15.2 (2)	<.001
Radiographic imaging	75 (75)	30 (40)		
Safe angulation	20 (20)	41 (54.7)		
Others	5 (5)	4 (5.3)		
Loading protocol			10.8 (1)	.001
Immediate	80 (80)	30 (40)		
Delayed (4-6 wk)	20 (20)	45 (60)		
Frequency of TAD[a] use			20.5 (1)	<.001
Rarely or sometimes	70 (70)	15 (20)		
Often or routinely	30 (30)	60 (80)		
Clinical applications			1.8 (1)	.18
Anterior retraction	50 (50)	36 (48)		
Others	50 (50)	39 (52)		
Operator preference			0.7 (1)	.41
Orthodontist	80 (80)	65 (86.7)		
Others	20 (20)	10 (13.3)		
Anchorage preference			0.5 (1)	.62
Direct anchorage	55 (55)	36 (48)		
Indirect or hybrid	45 (45)	39 (52)		

[a]TAD: temporary anchorage device.

Perception Results

Significant differences in perception-related items were observed between postgraduate trainees and faculty members. Postgraduate participants showed a stronger preference for radiographic imaging as the safest approach for mini screw placement, with 75% (75/100) selecting this option, whereas 55% (41/75) of faculty members predominantly favored safe insertion angulation techniques (*P*<.001).

Perceptions regarding loading protocols also varied notably: immediate loading was preferred by 80% (80/100) of postgraduate trainees, whereas delayed loading after 4 to 6 weeks was favored by 60% (45/75) of faculty members (*P*=.001). With respect to frequency of use, 70% (70/100) of postgraduate students reported rarely or occasionally using mini screws, whereas 80% (60/75) of faculty members indicated routine use (*P*<.001).

Despite these differences, perceptions related to clinical operator preference and anchorage type showed no statistically significant variation between groups, indicating general agreement in these areas. A detailed comparison of perception-related responses is shown in Table 2.

XSL•FO
RenderX

Attitude Results

Attitude-related findings indicated generally positive views on the clinical use of mini screws among both postgraduate trainees and faculty members. Operator preference was consistent across groups, with 80% (80/100) of postgraduate students and 87% (65/75) of faculty members indicating that mini screw placement should be performed by orthodontists, showing no statistically significant differences.

Similarly, attitudes toward anchorage type exhibited comparable patterns. Direct anchorage was selected by 55% (55/100) of postgraduate students and 48% (36/75) of faculty members, whereas indirect or hybrid anchorage approaches were reported at nearly equal rates across both groups, suggesting shared clinical attitudes toward anchorage selection.

In addition, several barriers influencing clinicians' attitudes toward broader adoption of mini screws were highlighted. These included concerns related to procedural safety, operator experience, and treatment cost, all of which contributed to variations in clinicians' readiness and confidence to integrate mini screws into routine orthodontic practice.

KAP Score Comparison

Table 3 shows that the mean KAP scores for male and female participants did not vary with any statistically significant difference. However, postgraduate students recorded significantly higher mean knowledge and perception scores when compared to faculty members; the attitudes between both groups were comparable. Among faculty, senior lecturers obtained somewhat higher scores than professors in perception, although this difference fell short of being statistically significant. When experience was considered, it was found that those participants who had less than 5 years of practice obtained the highest perception scores (mean 3.10, SD 0.85), which were significantly different from those of individuals with more than 10 years of experience (mean 2.62, SD 0.90). With respect to age groups, while the youngest group—26 to 35 years—attained the maximum knowledge and perception scores, the lowest mean scores were obtained in the age group of 46 to 50 years. Post hoc analysis further revealed statistically significant differences in attitude scores between the age group of 26 to 35 years and both the age groups of 36 to 45 years and 46 to 50 years.

Table 3. Mean knowledge, perception, and attitude scores with respect to age, gender, academic designation, and years of experience.

Variable and category	Number of participants	Knowledge score (–8 to +8), mean (SD)	*P* value	Perception score (–6 to +6), mean (SD)	*P* value	Attitude score (–6 to +6), mean (SD)	*P* value
Sex			.81		.65		.51
Male	95	3.15 (0.70)		2.87 (0.66)		1.92 (0.57)	
Female	80	3.20 (0.68)		2.79 (0.62)		2.05 (0.59)	
Academic designation			.01[a]		.02[a]		.73
Postgraduate students	100	3.66 (0.49)		3.29 (0.60)		2.04 (0.48)	
Faculty members	75	3.16 (0.48)	.15	2.39 (0.58)	.14	1.98 (0.50)	.63
Professors	20	3.10 (0.50)		2.35 (0.52)		1.92 (0.46)	
Senior lecturers	30	3.25 (0.47)		2.55 (0.53)		2.02 (0.44)	
Readers	25	3.18 (0.52)		2.45 (0.55)		2.00 (0.48)	
Experience (y)			.04[a]		.01[a]		.48
<5	62	3.68 (0.47)		3.10 (0.85)		1.95 (0.71)	
5-10	50	3.42 (0.53)		2.75 (0.82)		2.12 (0.52)	
>10	63	3.18 (0.55)		2.62 (0.90)		2.05 (0.60)	
Age (y)			.01[a]		.02[a]		.02[a]
26-35	85	3.43 (0.48)		3.05 (0.62)		1.96 (0.58)	
36-40	38	3.12 (0.51)		2.75 (0.65)		2.22 (0.57)	
41-45	22	2.95 (0.55)		2.62 (0.68)		2.04 (0.53)	
46-50	15	2.90 (0.60)		2.10 (0.70)		1.87 (0.54)	
51-55	15	3.43 (0.48)		2.83 (0.64)		2.16 (0.42)	

[a]Considered significant at $P<.05$.

XSL•FO
RenderX

Correlation Between KAP and Demographic Variables of the Participants

Table 4 shows that the KAP scores of participants varied negatively though weakly with age, years of experience, and designation. In other words, older participants, those with more than 10 years of clinical experience, and professors obtained relatively lower scores than younger individuals, those with fewer years of experience, and postgraduate students. These differences were statistically significant, particularly for perception scores. However, on the other hand, designation and years of experience showed a strong positive correlation with age. This means that, as the participants age, they hold higher academic positions and have a longer duration of professional practice. Most importantly, the correlations among knowledge, perception, and attitude scores were strongly positive; as knowledge improved, the perception and attitudes regarding the use of temporary anchorage devices (TADs) in orthodontics also improved.

Table 4. Correlation between knowledge, attitude, and perception scores and age, academic designation, and years of experience of the participants using the Spearman correlation coefficient test.

Variables	Age	Knowledge	Perception	Attitude	Academic designation	Experience
Age						
r	1	−0.12	−0.15[a]	−0.08	0.42[b]	0.55[c]
P value	—[d]	.09	.04[e]	.19	<.001[e]	<.001[e]
Knowledge score						
r	−0.12	1	0.74[c]	0.46[b]	−0.10	−0.18[a]
P value	.09	—	<.001[e]	.004[e]	.24	.03[e]
Perception score						
r	−0.15[a]	0.74[c]	1	0.51[b]	−0.14[a]	−0.20[a]
P value	.04[e]	<.001[e]	—	.002[e]	.03[e]	.02[e]
Attitude score						
r	−0.08	0.46[b]	0.51[b]	1	−0.11	−0.09
P value	.19	.004[e]	.002[e]	—	.21	.16
Academic designation						
r	0.42[b]	−0.10	−0.14[a]	−0.11	1	0.51[c]
P value	<.001[e]	.24	.03[e]	.21	—	<.001[e]
Experience						
r	0.55[c]	−0.18[a]	−0.20[a]	−0.09	0.51[c]	1
P value	<.001[e]	.03[e]	.02[e]	.16	<.001[e]	—

[a]Negative *r* values indicate a negative correlation (*r*=0.10-0.29 indicates a weak correlation).

[b]*r*=0.30 to 0.49 indicates a moderate correlation.

[c]*r*≥0.50 indicates a strong correlation.

[d]Not applicable.

[e]Considered significant at *P*<.05.

Discussion

Principal Findings

This study found that postgraduate trainees exhibited significantly higher knowledge and perception scores than faculty members, whereas attitudes were generally positive across both groups. Younger participants and those with fewer years of experience also showed more favorable views on mini screw use. This study also offers valuable insights into the awareness, perceptions, and clinical perspectives surrounding the use of orthodontic mini screws (TADs) among postgraduate trainees and academic faculty in the Kurdistan Region of Iraq. The findings reflect a pattern consistent with that found in the global literature, highlighting both notable progress and persisting challenges in integrating mini screws into routine orthodontic practice.

The comparative analysis revealed significant differences in foundational knowledge and preferred clinical protocols between postgraduate trainees and faculty members. Postgraduate trainees achieved higher knowledge and perception scores and tended to favor immediate loading, radiographic evaluation for safety, and direct anchorage approaches. In contrast, faculty members preferred delayed loading and placed greater emphasis on safe insertion angulation. These variations have been well

XSL•FO
RenderX

documented in previous research. For instance, Panaite et al [5] reported that both clinical experience and structured postgraduate training strongly influence the successful adoption and clinical mastery of mini screws. Nonetheless, the absence of consistent, hands-on education remains a global barrier to optimal application. Similarly, Al-Hammadi [13] identified insufficient training as a primary obstacle to broader adoption across the Middle East and Asia, a challenge that extends beyond Iraq.

The barriers identified in this study (operator experience, placement safety, and procedural cost) align closely with findings from China, Saudi Arabia, Canada, and Europe. A persistent training gap, compounded by perceived procedural complications, continues to slow widespread clinical integration. Reviews by Ahmed et al [14] and Panaite et al [5] further highlighted concerns regarding fracture resistance, optimal insertion protocols, and management of mini screw failure. However, recent evidence emphasizes that, when clinicians are trained through evidence-based protocols, mini screws consistently deliver predictable mechanical retention across a range of orthodontic scenarios [15,16].

The data from this study revealed a strong positive correlation among knowledge, perception, and attitude scores. In other words, as educational exposure increased, so did awareness, confidence, and willingness to incorporate mini screws into clinical practice [17]. Younger clinicians generally exhibited higher perception scores, reflecting their exposure to updated curricula and modern training environments. Senior faculty members, on the other hand, appeared more cautious, possibly due to established clinical routines and limited early exposure to TAD technology. This generational contrast underscores the importance of structured, continuous education at all professional levels [18].

Table 1 shows balanced participation between postgraduate trainees and faculty members, as well as between male and female respondents. Most (108/175, 61.7%) were aged 26 to 35 years, and nearly two-thirds (117/175, 66.9%) reported less than 5 years of professional experience. This demographic profile is consistent with similar regional and international studies, supporting the representativeness and reliability of this sample [5].

Table 2 compares the knowledge and clinical perceptions of postgraduate students and faculty members. Postgraduate students were more likely to attribute primary stability to bone quality, whereas faculty members emphasized cortical engagement—reflecting distinct experiential perspectives. Interestingly, faculty members reported higher perceived success rates (>90%), whereas postgraduate students estimated success within the 80% to 90% range. Statistically significant differences were found in preferred indications, safety protocols, loading strategies, and frequency of TAD use ($P<.05$). These outcomes align with existing literature indicating that accumulated experience often leads clinicians to adopt more conservative yet evidence-based techniques [19]. The general agreement between both groups regarding operator responsibility and anchorage principles suggests a shared understanding of fundamental clinical concepts.

Table 3 outlines the mean KAP scores across demographic variables. Gender was not a significant factor in any domain. Postgraduate students demonstrated significantly higher knowledge and perception scores than faculty members, although attitude scores were comparable between groups. Participants with less than 5 years of experience achieved the highest perception scores, whereas those with more than 10 years of practice recorded the lowest perception scores. Younger professionals (aged 26-35 years) also scored higher in both knowledge and perception. This trend aligns with international evidence indicating that early-career practitioners often exhibit greater enthusiasm and openness toward new technologies, including TADs, largely due to their exposure to evolving curricula and training systems [20].

Table 4 presents the correlations between demographic factors and KAP scores. Both age and years of experience exhibited weak but negative correlations with knowledge, perception, and attitude scores, meaning that older and more experienced clinicians tended to score lower in these areas. While this may seem counterintuitive, it is consistent with literature showing that senior practitioners often resist altering long-standing clinical approaches [21]. Nevertheless, the strong positive interrelationships among knowledge, perception, and attitude scores underscore that enhancing education and training can meaningfully improve clinician confidence and acceptance of mini screw use—an observation echoed across global research.

Overall, older participants, those with more than 10 years of professional experience, or those with senior academic ranks showed lower confidence, knowledge, and perceptions regarding mini screws than their younger counterparts. This generational divide likely stems from differences in educational exposure: senior clinicians were trained when skeletal anchorage concepts were not yet standard, whereas newer graduates have benefited from curricula and workshops incorporating TADs. Consequently, senior practitioners may be less inclined to modify established routines.

At the same time, differences in technological familiarity also shape these perspectives. Younger clinicians, more accustomed to digital resources and modernized learning platforms, tend to approach innovations such as mini screws with greater openness and confidence [21]. Encouragingly, many senior respondents expressed an interest in updating their skills and applying mini screw techniques in clinical practice. This willingness to engage in professional development offers a promising opportunity for improved uptake—particularly if supported through structured continuing education programs, mentorship models, and hands-on workshops tailored to different experience levels.

The findings of this study highlight a clear need to strengthen the educational infrastructure supporting the use of mini screws in orthodontic training and clinical practice. Integrating structured, evidence-based modules on TAD biomechanics, placement protocols, and complication management into postgraduate curricula could substantially improve clinician confidence and treatment outcomes. Additionally, tailored continuing education programs aimed at midcareer and senior orthodontists would help bridge the generational gap in knowledge and perception identified in this study.

XSL•FO
RenderX

Future research should focus on longitudinal assessments of how enhanced training and exposure influence clinical adoption and success rates of mini screws over time. Comparative studies across different institutions and regions in Iraq and broader Middle Eastern contexts could further elucidate contextual barriers to and enablers of technology uptake. Finally, mixed methods designs incorporating qualitative interviews may provide richer insights into clinicians' motivations, hesitations, and learning experiences related to skeletal anchorage systems. Such efforts will not only advance clinical practice but also contribute to the global dialogue on innovation adoption and educational reform in orthodontics.

Limitations

A review of the limitations of this study must include a note about the possibility of age bias. More recent concepts discussed in the questionnaire, such as newer techniques related to the use of mini screws, might not be known or attractive to older clinicians. Thus, the online mode of survey distribution (Typeform and other electronic platforms) may have contributed to sampling bias by favoring younger participants, who are more active on digital platforms. Our electronic survey tool, and specifically its implementation through Typeform, probably also favored more active younger professionals on digital platforms, hence causing an overrepresentation of early-career respondents.

The questionnaire was made in English as it is the language normally used for dental education and practice in this region. This posed a language barrier to those not very comfortable with the use of English, which could also mean that some people were left out. Likert scales were applied to closed-ended questions; although very easy to apply when collecting quantitative data, they may restrict the expression of different thoughts or detailed logic by respondents, hence introducing the possibility of misreading or oversimplification of varied attitudes and experiences. Reliance on electronic distribution platforms (email and professional WhatsApp groups) may have contributed to sampling bias by increasing participation among younger, technology-oriented clinicians.

Recommendations

This study reveals that orthodontists have a positive perception of mini screws and want to use them more often. To break the barriers of limited experience and safety, it is proposed to introduce focused training on mini screws in the dental education system and carry out targeted workshops for already established clinicians. This may be implemented through annual regional hands-on workshops, continuing professional development courses, and collaborative training programs involving dental colleges, teaching hospitals, and professional orthodontic associations. This will improve the skills and confidence in the safe adoption of mini screw techniques at all stages of a career.

Conclusions

Most participants appreciated the clinical advantages of mini screws and were ready to accept them as potential good anchorage tools. This study inferred that improving the availability of advanced technical resources in dental clinics accompanied by oriented education at the undergraduate and postgraduate levels can play a role in solving the existing hindrances against the wider application of mini screw techniques in orthodontic practice.

Funding

No external funding was received for this study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SMH contributed to conceptualization, data curation, formal analysis, project administration, methodology, and writing—review and editing. AMR contributed to conceptualization, survey design and distribution, supervision, and writing—review and editing.

Conflicts of Interest

None declared.

References

1. Gainsforth BL, Higley LB. A study of orthodontic anchorage possibilities in basal bone. Am J Orthod Oral Surg 1945 Aug:406-417 [FREE Full text] [doi: 10.1016/0096-6347(45)90025-1]
2. Atyah SI. Advances in orthodontic anchorage with the use of mini-implant techniques. The College of Dentistry, University of Baghdad. 2023. URL: https://codental.uobaghdad.edu.iq/wp-content/uploads/sites/14/2023/12/Sarah-Ismael.pdf [accessed 2026-01-28]
3. Kanomi R. Mini-implant for orthodontic anchorage. J Clin Orthod 1997 Nov;31(11):763-767. [Medline: 9511584]
4. Sharma R, Soni DM. Temporary anchorage devices (TADs) in orthodontics: a review. Asian J Res Med Med Sci 2023;5(1):1-11 [FREE Full text]

XSL•FO
RenderX

5. Panaite T, Balcos C, Savin C, Olteanu ND, Karvelas N, Romanec C, et al. Exploring the use, perceptions, and challenges of mini-implants in orthodontic practice: a survey study. Front Oral Health 2024 Jan 7;5:1483068 [FREE Full text] [doi: 10.3389/froh.2024.1483068] [Medline: 39844864]
6. Bungău TC, Moca AE, Ciavoi G, Romanul IM, Vaida LL, Buha CL. Usage and preferences of orthodontic mini-implants among Romanian practitioners: a survey study. Dent J (Basel) 2024 Dec 06;12(12):400 [FREE Full text] [doi: 10.3390/dj12120400] [Medline: 39727457]
7. Kyung HM, Ly NT, Hong M. Orthodontic skeletal anchorage: up-to-date review. Orthod Waves 2019 Nov 28;76(3):123-132 [FREE Full text] [doi: 10.1016/j.odw.2017.06.002]
8. Fritea L, Fodor R, Bei MF, Hackenberg B. Clinical utilization and practitioner preferences of orthodontic mini-implants in Romania: survey-based insights. Asian J Periodontics Orthod 2025;5(1):18-32 [FREE Full text] [doi: 10.51847/fvnu57lTFy]
9. Tarigan SH, Sufarnap E, Bahirrah S. The orthodontic mini-implants failures based on patient outcomes: systematic review. Eur J Dent 2024 May;18(2):417-429 [FREE Full text] [doi: 10.1055/s-0043-1772249] [Medline: 37848073]
10. Hsin-Chung Cheng J, De-Shing Chen D, Tan Y, Hu HT. Factors associated with usage frequency and pricing of temporary anchorage devices among orthodontists. J Dent Sci 2024 Jan;19(1):404-410 [FREE Full text] [doi: 10.1016/j.jds.2023.07.033] [Medline: 38303874]
11. Sample size calculator. Calculator.net. URL: https://www.calculator.net/sample-size-calculator.html [accessed 2026-01-28]
12. Murali S, Bagewadi A, Kumar S L, Fernandes A, Jayapriya T, Panwar A, et al. Knowledge, attitude, and perception of dentists regarding the role of artificial intelligence and its applications in oral medicine and radiology: a cross sectional study. J Oral Med Oral Surg 2023 Jul 10;29(2):22 [FREE Full text] [doi: 10.1051/mbcb/2023018]
13. Al-Hammadi MA. Current trends and challenges in the management of foot and mouth disease in Saudi Arabia: a review. Open Vet J 2025 May;15(5):1907-1933. [doi: 10.5455/OVJ.2025.v15.i5.6] [Medline: 40557100]
14. Ahmed N, Joseph R, Younus AA, Bhat KR. Temporary anchorage devices in orthodontics: a review. IP Indian J Orthod Dentofac Res 2020;6(4):222-228 [FREE Full text] [doi: 10.18231/j.ijodr.2020.044]
15. Truong VM, Kim S, Kim J, Lee JW, Park YS. Revisiting the complications of orthodontic miniscrew. Biomed Res Int 2022 Aug;2022(1):8720412 [FREE Full text] [doi: 10.1155/2022/8720412] [Medline: 35958810]
16. Alzahrani K, Alnafisah AM, Alzahrani KG, Alharbi KG. Knowledge, attitudes and practice of dental implants among dental interns in Saudi Arabia – a cross-sectional study. Open Dent J 2020 Sep 22;14:440-449 [FREE Full text] [doi: 10.2174/1874210602014010440]
17. Ashton KY, Jiang SS, Melo MA, Bosio JA. International investigation on temporary anchorage device use: a survey of orthodontists. J World Fed Orthod 2023 Jun;12(3):93-104 [FREE Full text] [doi: 10.1016/j.ejwf.2023.04.003] [Medline: 37198009]
18. Hyde JD, King GJ, Greenlee GM, Spiekerman C, Huang GJ. Survey of orthodontists' attitudes and experiences regarding miniscrew implants. J Clin Orthod 2010 Aug;44(8):481-486 [FREE Full text] [Medline: 21105585]
19. McGhee FC. Surgical technology faculty perceptions of using high-fidelity simulation in teaching. Walden University. 2025. URL: https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=19493&context=dissertations [accessed 2026-01-28]
20. Abu Al-Melh MM, Al-Anzi AN. Knowledge of undergraduate dental students toward orthodontic skeletal temporary anchorage devices at Kuwait University. BMC Med Educ 2020 Oct 06;20(1):340 [FREE Full text] [doi: 10.1186/s12909-020-02254-7] [Medline: 33023597]
21. Hussain SR, Jiang SS, Bosio JA. Generational perspectives of orthodontists in the U.S. and Canada: a survey study. Am J Orthod Dentofacial Orthop 2022 Dec;162(6):824-838. [doi: 10.1016/j.ajodo.2021.07.020] [Medline: 36055883]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
KAP: knowledge, attitudes, and perceptions
TAD: temporary anchorage device

Edited by E Mensah; submitted 31.Oct.2025; peer-reviewed by D Ali, H Othman; comments to author 10.Dec.2025; revised version received 12.Dec.2025; accepted 16.Jan.2026; published 26.Feb.2026.

Please cite as:
Hameed SM, Rauf AM
Knowledge, Attitudes, and Perceptions Regarding Mini Implants Among Orthodontic Practitioners in the Kurdistan Region: Cross-Sectional Study
Online J Public Health Inform 2026;18:e86896
URL: https://ojphi.jmir.org/2026/1/e86896
doi:10.2196/86896
PMID:41569888

XSL•FO
RenderX

©Soran M Hameed, Aras M Rauf. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 26.Feb.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparison of Artificial Intelligence Tools With Human Coding for Sentiment, Topic, and Thematic Analysis Tasks of Public Health Datasets During the COVID-19 Pandemic in Australia: Case Study

Danielle Hutchinson[1], PhD; Lauren Lee[1], MD; Haley Stone[1], PhD; Aye Moa[1], PhD; Holly Seale[2], PhD; C Raina MacIntyre[1], PhD

[1]Biosecurity Program, Kirby Institute, Faculty of Medicine and Health, UNSW Sydney, Wallace Wurth Building, Sydney, Australia

[2]School of Population Health, Faculty of Medicine and Health, UNSW Sydney, Sydney, Australia

Corresponding Author:
Danielle Hutchinson, PhD
Biosecurity Program, Kirby Institute, Faculty of Medicine and Health, UNSW Sydney, Wallace Wurth Building, Sydney, Australia

Abstract

Background: Public opinion, which may be influenced by personal experiences, news, and social media, can impact compliance with public health measures (PHMs) during health emergencies. Artificial intelligence (AI) tools offer opportunities to analyze public opinion in real time during health emergencies. However, their performance in accurately identifying sentiment and themes in health-related online content remains unclear.

Objective: This study aimed to evaluate the performance of natural language processing–based and large language model (LLM)–based AI tools when compared to human coding for sentiment analysis, topic modeling, and thematic analysis of public health datasets. Tools were selected to reflect those available to public health analysts and decision-makers.

Methods: Data were collected via Google Alerts (GA) and social media posts from X (formerly known as Twitter) relevant to COVID-19 mitigation PHMs from December 2022 to February 2023. Following relevance screening, the sentiment of the complete datasets was analyzed by a human rater, with descriptive statistics used to summarize the overall sentiment profile. Subsets of 400 GA articles and 400 tweets were manually coded for sentiment by 2 human raters. Results were compared with outputs from 5 AI tools, including VADER (Valence Aware Dictionary and Sentiment Reasoner), SentimentGI, SentimentQDAP, Microsoft Azure, and OpenAI's ChatGPT-4. Topic modeling of the GA and X datasets was conducted using latent Dirichlet allocation in R and zero-shot prompting in ChatGPT-4 and compared with manual topic summaries. Thematic analysis of positive and negative sentiment datasets was conducted by a human rater and ChatGPT-4, with outputs evaluated for proficiency and reasonableness. The sentiment of the entire datasets was analyzed by a human rater, and descriptive statistics were calculated.

Results: Of 2227 GA results and 3484 tweets, 58% (n=1238) and 71% (n=2473), respectively, were relevant to PHMs. Human-coded sentiment analysis showed mostly neutral reporting in the news media, while social media expressed more polarized views. Across both datasets, AI tools demonstrated poor concordance with human-coded sentiment (Cohen κ <0.5 for all tools and sentiment categories). Topic modeling with ChatGPT-4 aligned more closely with human-rated topics than latent Dirichlet allocation, and of the 20 LLM-generated thematic outputs, 13 were rated proficient, and 7 were rated partially proficient. LLM outputs provided coherent, high-level summaries but lacked contextual insight. Human and LLM thematic analyses both identified themes of vaccine effectiveness, debate regarding PHMs, and public trust.

Conclusions: Accessible AI tools demonstrate limited reliability for sentiment classification of health-related online text but show promise for rapid thematic exploration when combined with human oversight. These tools could complement traditional qualitative research in the context of health emergencies; however, they require human review to enhance the accuracy of interpretation. Further research is needed for non-English datasets.

(Online J Public Health Inform 2026;18:e80824) doi:10.2196/80824

KEYWORDS

public opinion; sentiment analysis; social media; artificial intelligence; AI; public health informatics; equity; COVID-19

XSL•FO
RenderX

Introduction

In the context of health, public opinions can change over time, vary across the population, and are often influenced by factors such as personal experiences and media exposure [1]. Public opinion can impact the course of an epidemic through impacting levels of compliance with public health measures (PHMs) such as vaccines, mask-wearing, and social distancing [2]. For example, Yu et al [3] used agent-based modeling to describe the relationship between the spread of COVID-19 and opinion dynamics in 15 different countries and found that public opinion on preventive interventions impacted the cumulative number of cases, particularly in the early stages of an epidemic. The World Health Organization has identified disinformation, the intentional spread of misleading information, as a threat to public health [4]. This may occur by changing the opinions and, therefore, the behaviors of populations through the creation of uncertainty about PHMs [4,5]. Therefore, the collection of public opinion data is important to measure public acceptance of PHMs and to monitor changes over time, thereby serving as a tool to combat the impact of disinformation and promote compliance with public health advice. There has been a proliferation of published research using artificial intelligence (AI) tools to analyze these data [6]. However, there is limited understanding of the accuracy of these readily available AI tools in conducting sentiment, topic, or thematic analysis on datasets of public opinion in real-world scenarios.

Public opinion data provide an important feedback mechanism during health emergencies such as the COVID-19 pandemic [7]. Qualitative research methods, such as analysis of data collected from focus groups and individual interviews, are recommended to better understand community opinions about disease perception and preventive behaviors to inform response efforts during health emergencies [8-10]. However, barriers to rapid qualitative research in the context of an infectious disease outbreak may include difficulties in conducting focus groups and interviews due to exposure risks, and participants may be reluctant to participate in the study due to the impact of the disease or the public health response [8,10]. In emergency situations, it is important to share findings in almost real time; therefore, the time taken to conduct qualitative research is a barrier, particularly during health emergencies [8,10]. Rapid data analysis techniques may deliver some time savings; however, data collection, interpretation, and write-up of results remain time intensive [11], with the World Health Organization European guidelines suggesting that the entire process takes 4 to 6 weeks, and Dong et al [12] classified rapid qualitative methods as those taking less than 6 months from conception to reporting of results [9,12]. Comparatively, data collection and analysis using AI can take as little as a few minutes [13]. Other barriers may include the capacity to form a research team of available staff with the necessary expertise to undertake qualitative analysis of community opinion to inform health policy [8].

Health emergencies such as the COVID-19 pandemic have led to the consideration of new methodologies for the collection and use of evidence to inform policy decision-making [14]. Previous studies have proposed the use of AI tools, such as natural language processing (NLP) and large language models (LLMs), to reduce the workload and provide real-time insights to help inform public health decision-making [15-19]. AI tools can be used to complement traditional qualitative data analysis methods in public health through sentiment analysis, topic modeling, and thematic analysis and have been shown to be less time-consuming and resource-intensive for these tasks [13,20-23]. Access to platforms to perform social media analysis, as well as improved training and capacity to conduct this type of research, was identified as key areas for improvement in a global survey of public health professionals [24]. Our study has specifically chosen tools that are readily available and do not require in-depth training for a public health professional to use.

Despite the proliferation of published research on public opinion data, collected via surveys and social media, toward PHMs during the COVID-19 pandemic, it is unclear whether these data are used as part of evidence-based policy decision-making [25].

This study adopted an applied informatics perspective, focusing on tools and data pipelines readily available to Australian public health analysts. The study aimed to evaluate the feasibility and performance of commonly available AI text analysis tools in analyzing public health datasets from social media (X) and news media (Google Alerts [GA]). Specifically, the study examined how traditional NLP methods and LLM tools perform when analyzing online public opinion data about PHMs during the COVID-19 pandemic in Australia.

The study addressed two research questions:

1. How accurately and consistently do AI-based tools classify sentiment and extract relevant topic and thematic content from public health–related online text?
2. What role does human oversight play in ensuring the validity and reliability of AI-driven sentiment and thematic analysis in public health informatics workflows?

First, we hypothesized that AI tools would demonstrate generally limited accuracy for broad sentiment classification and that LLMs would perform better than NLP-based text analysis tools in identifying and summarizing thematic content. Second, we hypothesized that human review and interpretive oversight will remain necessary to ensure data quality and contextual understanding of online public health datasets, underscoring the continuing importance of human oversight in public health informatics workflows.

Methods

Data Sources and Sampling

This study used open-source online data to evaluate the feasibility of automated methods for analyzing public opinion about PHMs during the COVID-19 pandemic in Australia. Two platforms were selected: GA and X (formerly known as Twitter), representing traditional and social media sources, respectively, and the most commonly used sources in infodemiology research [26].

XSL•FO
RenderX

GA is a change detection and notification system that automatically monitors multiple websites for mentions of a textual string and allows the user to select the frequency of monitoring, the source, language, and region [27,28]. GA was configured to retrieve Australian news articles that included key COVID-19–related terms (Table 1) once daily [28]. Each alert provided the publication date, publisher, article title, and 2-line summary, which were collected and stored in a Microsoft Excel spreadsheet. Data collection occurred from December 19, 2022, to February 19, 2023, during the fourth Omicron COVID-19 wave in Australia [29].

Table . Search terms.

Public health measure	Search terms
Vaccination	Vaccine, booster, vax, jab, Pfizer, Moderna, Novavax, (kids OR children) AND (Vaccine)
Mask-wearing	Mask, n95 OR respirator, school AND masks, (kids or children) AND (mask or unmask), public transport AND mask
Lockdown	Lockdown
Isolation requirements	Isolation, iso
Mandates	Mandate
Ventilation	Ventilation

The social media platform X is one of the most widely used platforms for health research and was freely available with an academic license at the time of data collection [30,31].

A Python app was used to access the X academic application programming interface (API) and to search for tweets geolocated in Australia using identical keyword combinations and date ranges. Specifically, we used a custom script to handle query construction, data collection, and filtering by keywords, language, and location. Separate searches were conducted for each keyword. All English-language tweets posted in Australia on these days were retrieved and included in the relevance assessment. Tweets with geolocation outside Australia and non–English-language tweets were excluded. Tweet ID and date of publication, tweet content, region, and location were collected and stored in a Microsoft Excel spreadsheet. User information was not collected to maintain anonymity.

A random sample of 800 items (400 GA articles and 400 tweets) was drawn using the RAND() function in Microsoft Excel (version 2306; Microsoft 365) to produce a manageable dataset for manual validation and automated analysis [32]. This dataset size allowed comparison across multiple sentiment and thematic analysis tools while maintaining a feasible human coding time for the manually coded reference set.

Data Cleaning and Preprocessing

Text data from both platforms were preprocessed to standardize the format and remove noise before analysis. Duplicates were removed from the GA dataset by checking for duplicate URLs, thereby retaining articles with similar content published across different days or in different publications (eg, syndicated articles). Within the X dataset, duplicate tweets were removed, while quote tweets and retweets were preserved for analysis. This allowed for analysis of the volume of public interest across news media and social media regarding the topics of interest. Data preprocessing included the removal of Twitter handles, URLs, stop words, and punctuation. Individual GA articles and tweets that discussed more than one PHM (eg, both mask-wearing and vaccination) were disaggregated and analyzed separately for each relevant PHM category.

All collected GA results and tweets were assessed for relevance to PHMs as per the inclusion and exclusion criteria (Textbox 1) by 1 reviewer (DH). A subset of 100 tweets and 100 news articles was assessed by a second reviewer (HS), and interrater reliability was assessed using the Cohen κ coefficient [33].

Textbox 1. Inclusion and exclusion criteria for relevance of articles to public health measures.

Inclusion criteria

- Subjective (opinions, editorial, and commentary) or objective (information and policy announcements) discourse related to pharmacological (vaccines or other) and nonpharmacological (masks, ventilation, physical distancing, stay-at-home orders or lockdowns, and school closures) public health measures for communicable diseases (any).
- General information about vaccines

Exclusion criteria

- Scientific or medical journal papers or reports of the same
- Articles about case numbers or modeling
- Articles not relating to communicable disease
- Articles not relating to public health measures: pharmacological (vaccines or other) and nonpharmacological (masks, ventilation, physical distancing, stay-at-home orders or lockdowns, and school closures)

XSL•FO
RenderX

Human-Coded Sentiment Analysis

Human-coded sentiment analysis involves individuals manually reviewing textual data and assigning a sentiment label to each text based on its content [34]. A data subset comprising 400 GA and 400 tweets was randomly selected using the RAND() function in Microsoft Excel and reviewed by 2 reviewers (DH and LL). As only the first 2 sentences of the news article were included in the GA dataset, the search term was also provided to assist the reviewer to determine sentiment. Sentiment toward the PHM was assessed and assigned a positive, negative, or neutral label based on the interpretation of the text. If the sentiment toward the PHM was positive, but the sentiment of the whole tweet was negative, it was assigned a positive value.

The reviewers used a rule-based process with the development of annotation guidelines (Textbox 2) and met after the first 50 items to review the guidelines and discuss any clarifications needed. Reviewers met again at the completion of the analysis to address any ambiguous cases. Interrater reliability was calculated using the Cohen κ coefficient [33]. Disagreements between human raters were resolved through discussion and consensus to ensure consistency in the final labeled dataset used for AI comparison. One reviewer (DH) completed the sentiment annotation of the remaining relevant GA and X datasets. GA results and tweets that did not contain sufficient information to assign sentiment were removed from the datasets. GA results and tweets that discussed more than one PHM were analyzed separately for each relevant PHM category, allowing for assessment of sentiment associated with specific PHMs within the same source text. GA results and tweets labeled positive, negative, and neutral were aggregated for analysis. Search terms relating to vaccination and mask-wearing (Table 1) were combined for further analysis; for example, the results for "vax," "jab," and "vaccine" were collapsed under the umbrella term "vaccine." The proportion of sentiment, expressed as a percentage of the total number of GA results and tweets, was calculated with 95% CIs.

Textbox 2. Annotation guidelines for human-coded sentiment analysis of articles and tweets.

Positive sentiment

- Positive words referencing the public health measure
- Criticism of opinions against the public health measure (eg, criticism of antivaxxers was assigned a positive sentiment toward vaccines)

Neutral sentiment

- Neutral reporting of events and announcements
- Mention of public health measures without expressed sentiment or opinion

Negative Sentiment

- Negative words referencing the public health measure
- Criticism of opinions supporting the public health measure

AI Tools Evaluated for Sentiment Analysis Task

A total of 5 AI-based text analysis tools were evaluated to compare the performance of traditional NLP techniques and LLM-based methods, which were validated against human-coded reference data. Tools were selected based on their accessibility for public sector analysts and compatibility with existing Australian government data science infrastructure [35-38]. R (version 4.3.1; R Foundation for Statistical Computing) within RStudio (version 2023.06.1; Posit Software), sentiment analysis tools (Valence Aware Dictionary and Sentiment Reasoner [VADER], SentimentGI, and SentimentQDAP), Microsoft Azure Machine Learning, and ChatGPT-4 represent platforms that are free, low cost, or integrated within standard analytics environments used by public health agencies [39-43]. Each text entry was assigned a compound sentiment score ranging from −1 (most negative) to +1 (most positive), which was then categorized as positive, neutral, or negative using default thresholds (≥0.05 positive; ≤−0.05 negative).

NLP Tools in R

R is used extensively in public health research and includes packages specifically designed for cleaning and visualizing large public health datasets [35].

VADER is a lexicon- and rule-based sentiment analysis tool optimized for social media text. It was implemented in R using the vader package [44].

The SentimentAnalysis package (SentimentGI and SentimentQDAP) in R extends lexicon-based sentiment scoring by incorporating valence shifters (eg, negators and amplifiers) that adjust for linguistic nuance, which makes it particularly suited to conversational or social media–style text [45].

Microsoft Azure

Microsoft Azure Machine Learning Text Analytics (version 3.1; 2023) was used to assess sentiment through integration into the Microsoft Excel spreadsheet environment via the Power Query function, with results returned as sentiment probabilities for positive, neutral, and negative categories [46]. The Microsoft suite is widely used across Australian government agencies and was therefore included in this study [38]. API version and date of access were documented to ensure reproducibility.

XSL•FO
RenderX

OpenAI ChatGPT (GPT-4)

ChatGPT was tested as an emerging LLM approach for qualitative text classification. Using OpenAI's API (March 2024) [47], each text item was submitted with a structured zero-shot prompt (Table S1 in Multimedia Appendix 1). Responses were parsed programmatically for sentiment and theme. This approach reflects the real-world use of generative AI in rapid public health analysis while maintaining reproducibility through prompt documentation.

Topic and Thematic Analysis

To demonstrate how these data could be used in the public health domain for decision-making and evaluation, topic summary and thematic analyses were performed. Topic summary analysis was performed on the entire dataset by identifying key phrases and trends manually by highlighting recurring words, phrases, or patterns that may represent shared topics within the datasets [48]. For comparison, topic modeling was conducted on the GA and X datasets using latent Dirichlet allocation (LDA), a probabilistic modeling technique used to identify topics occurring in a textual dataset [22]. This was done using the "topicmodels" package in R [49], and the code is included in Multimedia Appendix 2. ChatGPT-4 was used to replicate the manual topic selection using a zero-shot prompting technique outlined in Table S1 in Multimedia Appendix 1 [50]. Results were collated, and the top 5 topics from each method were compared.

For the thematic analysis, the positive and negative results in the GA and X datasets were extracted from the entire annotated dataset and put into 4 separate files. Neutral sentiment results were excluded. A qualitative thematic analysis explored the opinions expressed toward PHMs within the Australian news media and among X users. This was done using Braun and Clarke's [51] 6-step process, coding each data line prior to identifying common themes across the dataset. Examples from the datasets were collated, with paraphrasing of tweets to maintain user anonymity. For comparison, ChatGPT-4 was used to replicate the manual thematic analysis [52] on the 4 datasets (GA positive, GA negative, X positive, and X negative results) using the zero-shot prompting technique outlined in Table S1 in Multimedia Appendix 1 [50]. The thematic analysis results were presented with subthemes and paraphrased representative sentences for each dataset.

Statistical Analysis

Concordance of each AI tool with human sentiment analysis was calculated by comparing the rating of a subset of tweets and articles with the human-rated sentiment analysis. Interrater reliability was calculated using the Cohen κ coefficient [33]. The κ statistic was used to assess interrater agreement, and results were classified as 0 to 0.5 weak, 0.51 to 0.8 moderate, and 0.81 and above strong [33]. Statistical analysis was performed using R, where a 2-sided P value of <.05 was considered statistically significant [49].

Validation and Comparison

All tool outputs were compared against the manually coded dataset. Descriptive and statistical analyses were performed on the results of the sentiment analysis process using Microsoft Excel [32]. For each sentiment (eg, "positive"), we identified all cases where the human rater assigned that sentiment and then calculated the percentage of those cases in which the AI tool assigned the same sentiment (Table S2 in Multimedia Appendix 1). This process was repeated for each sentiment category (positive, neutral, and negative) and each tool.

All statistical analyses were performed using R [53].

The accuracy of the generative AI results for topic modeling was calculated using the number of agreements in each dataset [52]. This was done using a cross-matching rubric (Table 2).

XSL•FO
RenderX

Table . Cross-matching rubric for alignment of human and artificial intelligence topic modeling and thematic analysis results.

Tasks and score	Label	Definition
Topic modeling		
2	Direct match	Strong alignment with any manually generated topic. Captures core meaning.
1	Partial match	Aligns with a related concept, but not the central idea. May be too generalized.
0	No match	Does not align with any manually generated topic.
Thematic analysis: proficiency		
—[a]	Proficient	The LLM-generated[b] theme matched closely and accurately with high relevance to the human coding.
—	Partially proficient	The theme matched moderately well, with some relevance to the human coding.
—	Not proficient	The LLM theme matched poorly with little relevance to the human coding.
Thematic analysis: reasonableness		
2	Very reasonable	The theme had high relevance to the dataset and would likely be generated by human coding.
1	Reasonable	The theme had some relevance to the dataset and could be generated by human coding.
0	Not reasonable	The LLM-generated theme had little relevance to the dataset.

[a]Not applicable.

[b]LLM: large language model.

A comparative analysis of the human and LLM thematic analyses was conducted by 1 author (DH), who reviewed each output of the LLM and assessed whether the theme matched the human thematic analysis. A cross-matching rubric, similar to the one used for topic analysis agreement, was developed to classify the LLM-generated output as “proficient,” ”partially proficient,” or ”not proficient” in capturing themes from the data that matched the human coding (Table 2) [50,52]. For all LLM outputs, regardless of whether they matched the human output, a score for how “reasonable” it was to derive the theme from the dataset was given, using a scale of 0 to 2, as has been used in similar studies [52]. The scoring was completed by 1 author, a subject matter expert (DH; Table 2).

Ethical Considerations

Nonidentifiable data from online news media and social media were collected during this study. We did not analyze individual accounts and have not published any identifiable information or individual quotes. The LLM used in this study, ChatGPT-4, was used in a strictly limited capacity for analyzing nonsensitive, deidentified text. The research team ensured that no identifiable data were shared with the AI platform, and all use complied with institutional research integrity and data privacy guidelines. All data were deidentified (including the removal of Twitter handles) prior to input, and chat history was disabled to prevent storage or reuse of content by the model provider. Chats were deleted when the session was complete. The study was approved by the UNSW Human Research Ethics Committee (approval number HC230028).

Results

Overview

The results of this study are presented as follows. First, a description of the dataset is given. Second, the results of the comparison of human-coded and machine-coded sentiment analysis on the data subset are outlined. Third, a summary of the sentiment analysis of the entire dataset is provided. Finally, human-generated and machine-generated topic modeling and thematic analysis results of the entire dataset are compared to demonstrate the feasibility of using AI tools to assist public health analysts in assessing public opinion during health emergencies.

Description of the Dataset

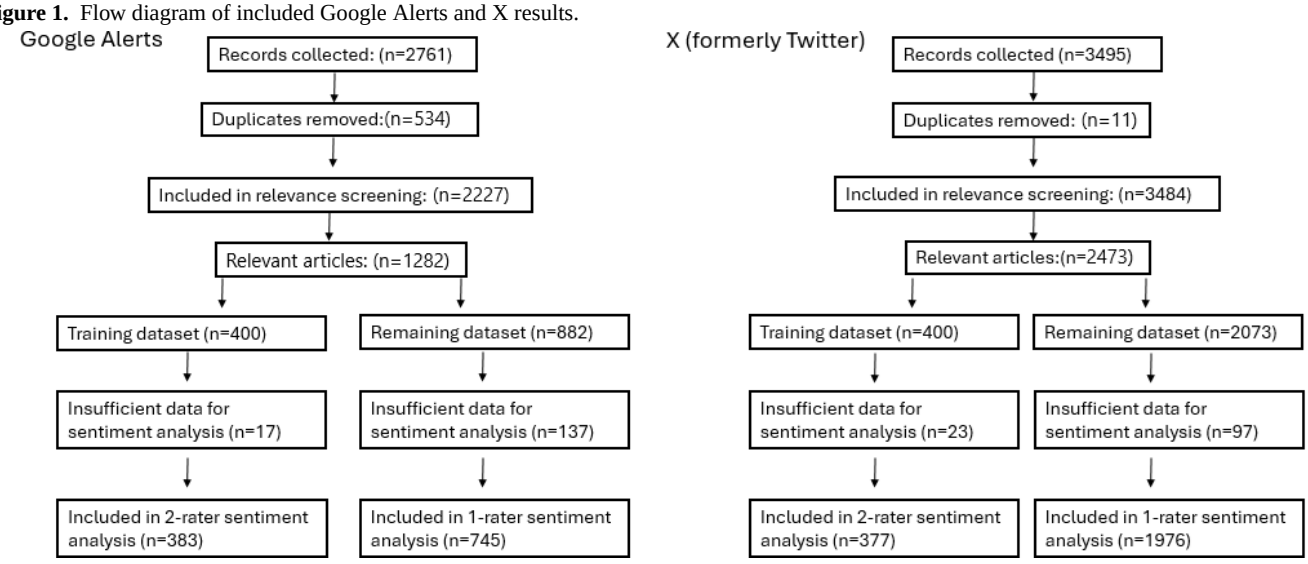
Overall, 2761 GA articles and 3495 tweets were collected during the study period. Following removal of duplicates, 2227 GA articles and 3484 tweets were included in the relevance screening, with 57.6% (n=1283) of GA articles and 71% (n=2473) of tweets assessed as relevant. There were moderate (GA: 88% concordance, κ=0.76) and strong (X: 94% concordance, κ=0.86) levels of agreement for relevance between 2 analysts in the data subsets of 100 (88%) GA articles and 100 (100%) tweets.

During the sentiment analysis process, further GA articles (n=154) and tweets (n=120) were removed from the dataset, as there was not enough information to determine sentiment (Figure

XSL•FO
RenderX

1). There were moderate levels of agreement for sentiment scores between the 2 reviewers in both the GA dataset (κ=0.69; *P*<.001) and the X dataset (κ=0.75; *P*<.001; Table S2 in Multimedia Appendix 1).

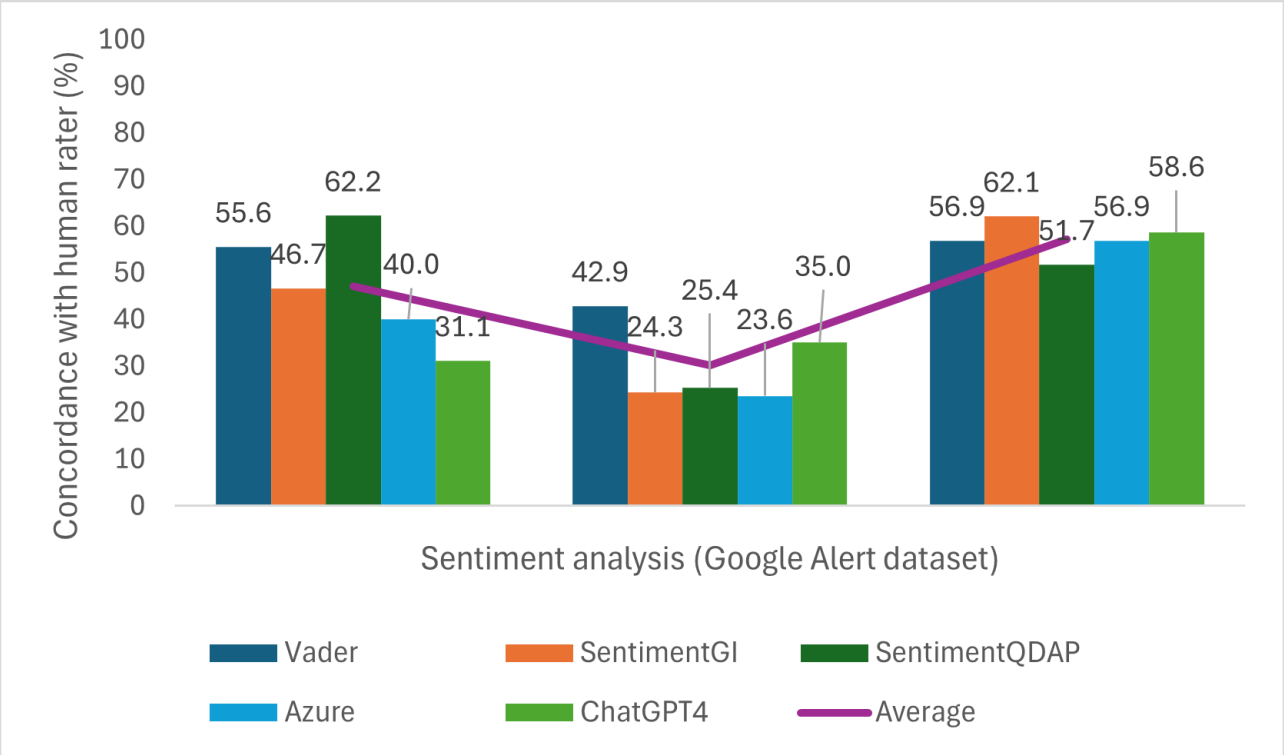
Figure 1. Flow diagram of included Google Alerts and X results.



Sentiment Classification Performance of AI Tools

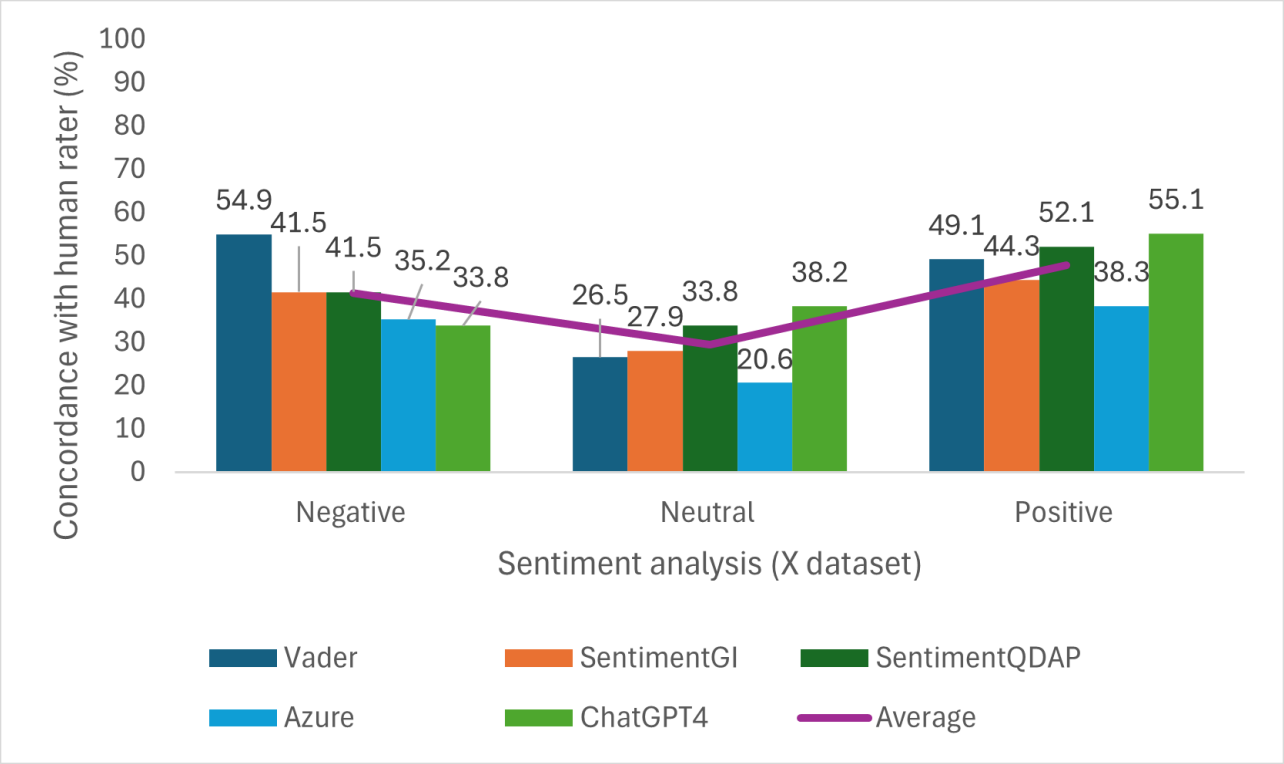
The results of sentiment classification performance are presented as proportion (%) of agreement with the human rater for each tool and sentiment category (Figures 2 and 3). The performance of the AI tools was generally low and varied across sentiment categories and data sources. Agreement was highest for positive sentiment and lowest for neutral sentiment across both platforms. In the GA dataset, SentimentQDAP had the highest agreement with the human rater for negative sentiment (n=28, 62.2%), while SentimentGI had the highest agreement for positive sentiment (n=36, 62.1%), and VADER performed best for neutral sentiment in this dataset, with an agreement rate of 42.9% (n=120). In the X dataset, VADER achieved the highest agreement for negative sentiment (n=78, 54.9%), while ChatGPT-4 demonstrated the highest agreement for both neutral (n=48, 38.2%) and positive (n=92, 55.1%) sentiments. An interrater reliability analysis was performed between the dependent samples of reviewer 1 and each AI tool. For this purpose, Cohen κ was calculated, and in all cases, no agreement was found (Table S3 in Multimedia Appendix 1).

Figure 2. Comparison of human-rated sentiment score for a subset of Google Alerts articles (n=383) with 5 artificial intelligence tools.



XSL•FO
RenderX

Figure 3. Comparison of human-rated sentiment score for a subset of tweets (n=377) with 5 artificial intelligence tools.



Manual Sentiment Analysis of Entire Datasets

Total Sentiments Across the Datasets

Positive, negative, and neutral sentiment results were collated (Table 3), with differences in the distribution of sentiment observed between the 2 platforms. Although the relevant datasets comprised 1283 (57.6%) GA articles and 2473 (%) tweets, a higher number of text segments were analyzed due to articles and tweets that referred to multiple PHMs. These were coded and analyzed separately for each PHM to measure sentiment patterns. In the analysis of the GA platform (n=1587), the majority of the sentiment was neutral (n=1021, 64.3%, 95% CI 62% - 66.7%). Negative sentiments accounted for 18.1% (n=287, 95% CI 16.2% - 20%), while positive sentiments accounted for 17.6% (n=279, 95% CI 15.7% - 19.5%). In the analysis of text segments on the X platform (n=3124), sentiment demonstrated a higher degree of polarity, with negative sentiments comprising 40% (n=1248; 95% CI 38.2% - 41.7%) and positive sentiments comprising 39.5% (n=1233; 95% CI 37.8% - 41.2%) of the total. In the X dataset, 20.6% (n=643; 95% CI 19.2% - 22%) of sentiments were neutral. These results indicate fewer opinions and more neutral reporting of events in the GA dataset, while there was a higher degree of both positive and negative emotional expression found within the X dataset.

Table . Sentiment distribution across platforms in the entire datasets (Google Alerts [GA] and X).

Platform and sentiment	Frequency (n)	Proportion (95% CI)
GA (n=1587)		
Negative	287	18.1% (16.2% - 20%)
Neutral	1021	64.3% (62% - 66.7%)
Positive	279	17.6% (15.7% - 19.5%)
X (n=3124)		
Negative	1248	40% (38.2% - 41.7%)
Neutral	643	20.6% (19.2% - 22%)
Positive	1233	39.5% (37.8% - 41.2%)

Sentiment Analysis of Each Search Term

Sentiment analysis results were collated for each search term in each dataset (Table S4 in Multimedia Appendix 1) to investigate sentiment expressed toward specific topics and specific PHM. As described earlier, the various PHMs were discussed in different tones across the 2 platforms, with GA results more likely to be classified as neutral sentiment, and X results more likely to show polarity for each search term. Neutral GA sentiments were expected due to the reporting of events in the news media. For search terms relating to pharmaceutical companies in the X dataset, reports on the search term

XSL•FO
RenderX

"Moderna" (n=30), sentiment was more evenly distributed (negative: n=11, 36.7%; neutral: n=10, 33.3%; and positive: n=9, 30%) when compared to sentiment for the search term "Pfizer" (n=222), which was mostly negative (n=135, 60.8%), 30.6% neutral (n=68), and 8.6% positive (n=19). When comparing the search term "booster" (n=187) in X dataset, sentiment was broken down as 40.1% (n=75) negative, 16% (n=30) neutral, and 43.9% (n=82) positive; however, sentiment for the search term "vaccine" (n=860) was mostly negative (n=378, 44%), 28.4% neutral (n=244), and 27.7% positive (n=238). Search terms relating to lockdown (n=23, 54.8%) and mandates (n=85, 66.4%) were associated with negative sentiment in the X dataset, while isolation was associated with positive sentiment (n=17, 85%).

Sentiment Analysis of Vaccination and Mask Search Terms

Results for search terms related to COVID-19 vaccination and mask-wearing were combined (Table S5 in Multimedia Appendix 1). Vaccine results were analyzed with and without the inclusion of results of the search term "booster," as it was shown to have opposing polarity when compared to "vaccine" (Table S5 in Multimedia Appendix 1). Analysis of the combined vaccine search terms continues to show the same pattern of distribution between the 2 datasets, with sentiment in the GA dataset predominantly neutral (65.8%) and lower proportions of negative (18.1%) and positive (16.1%) sentiment. In contrast, the X dataset was more polarized and critical of vaccines, with nearly half (49.5%) of all results expressing negative sentiment and lower levels of neutral (26.9%) and positive (23.6%) sentiment (Table S5 in Multimedia Appendix 1). When booster-related results were included in the analysis, these patterns were retained.

The pattern was markedly different for sentiment analysis results of reports collected via mask-related terms. The GA results were also mostly neutral (61.4%), and 26.4% expressing positive sentiment and only 12.3% expressing negative sentiment. In the X dataset, there was a much higher proportion of positive sentiment (66.8%) than negative (21.5%) or neutral (11.7%) sentiments expressed.

Topic Summary Analysis

LDA topic modeling of the GA and X datasets (Tables S6 and S7 in Multimedia Appendix 1) and the LLM results were compared to the human rating of the top 5 most discussed topics and are presented in Tables 4 and 5, including alignment scores from the cross-matching rubric (Table 2).

Table . Comparison of human-generated and machine-generated topic analysis of the Google Alerts dataset.

Manually generated topic	RStudio LDA[a] output	Machine score	LLM[b] output	LLM score
Prominent physician reveals vaccine injury	Public attitudes toward vaccines and masks	0[c]	Public debate on COVID-19 vaccines	1[d]
China allows travel after lifting of COVID-19 restrictions	COVID-19 mitigation: mask usage and vaccination	1	Mask mandates and usage	1
Impact of vaccine rollout on economy	COVID-19 protection and boosters	1	Economic and social impacts of PHM	2[e]
Easing of COVID-19 restrictions	Australian COVID-19 response	2	COVID-19 vaccine mandates and compliance	1
India—wearing masks due to new COVID-19 wave	Intentions regarding public health measures	0	Global or local responses to COVID-19 variants	2

[a]LDA: latent Dirichlet allocation.

[b]LLM: large language model.

[c]0: no match.

[d]1: partial match.

[e]2: direct match.

XSL•FO
RenderX

Table . Comparison of human-generated and machine-generated topic analysis of X dataset.

Manually generated topic	RStudio LDA[a] output	Machine score	LLM[b] output	LLM score
COVID-19 vaccines	Discourse on mask-wearing and COVID-19 vaccination	2[c]	Vaccine safety and side effects	1[d]
New isolation rules—5 days with mask	Individual attitudes toward masks and vaccination	2	Mask usage and effectiveness	2
Wearing masks on a plane	COVID-19 choices including vaccine brands	0[e]	Vaccination campaigns and compliance	2
Wearing masks on public transport	Intention to comply with PHM[f]	1	Public health measures and policy debate	2
China ceases zero-COVID-19 policy	Current perspectives on mask-wearing and vaccination	1	Misinformation and distrust in authorities	0

[a]LDA: latent Dirichlet allocation.

[b]LLM: large language model.

[c]2: direct match.

[d]1: partial match.

[e]0: no match.

[f]PHM: public health measure.

GA Dataset

Mask use and mask mandates, as well as vaccine mandates, are discussed in the GA dataset within the context of easing COVID-19 restrictions (Table 4). Across the 5 cases, the RStudio LDA output topics fully aligned with 1 of 5 cases and partially aligned with the manual coding in 2 of 5 instances, while the LLM output fully matched in 2 of 5 instances and partially matched in 3 of 5 cases. The LLM outputs show a deeper understanding of the context but are often too generalized, while some of the machine-generated outputs were related but not specific to the manually generated topics.

X Dataset

Topics discussed in the X dataset included announcements concerning the easing of restrictions, and opinions both supporting and opposing the dropping of mask mandates and prolonged isolation following a positive COVID-19 test (Table 5). Vaccination campaigns and safety concerns were also discussed. Results from the cross-matching rubric of the X dataset showed that the LLM output more closely aligned with the manually generated topic, fully matching in 3 of 5 cases and partially matching in 1 of 5 cases. The LDA output fully matched in 2 of 5 instances and partially matched in 2 of 5. The LLM output is more descriptive of the topics.

Human-Coded Thematic Analysis

Thematic analysis was performed on the positive- and negative-assigned GA articles and tweets (Tables S8a-S8d in Multimedia Appendix 1). Many tweets were of the account owners reporting their choice to engage or not engage with particular PHMs (eg, “My wife and I wear masks when we go out unless we are eating or drinking or outside. There is no way I am having that experimental vaccine”). Several other themes emerged and are summarized in Table 6.

XSL•FO
RenderX

Table . Summary of thematic analysis results by large language model (LLM).

Dataset	LLM summary
Positive GA[a] dataset	"The thematic analysis reveals a comprehensive approach to managing COVID-19, from vaccination campaigns and mask mandates to broader societal adjustments. Public health measures, supported by clear communication, have been instrumental in mitigating the impact of the pandemic. Emerging variants and evolving immunity continue to challenge and shape policies. Overall, the findings underscore the importance of ongoing vigilance and adaptability in public health."
Negative GA dataset	"The thematic analysis reveals ongoing tensions surrounding vaccine mandates, including public resistance and concerns over safety. Mask mandates evoke polarized opinions, with debates on their necessity and effectiveness. Lockdowns have left lasting societal and psychological impacts, fuelling protests and dissatisfaction. Public health messaging faces challenges due to misinformation and varying adherence levels. The emergence of new COVID-19 variants highlights the need for adaptive vaccination strategies. These findings underscore the complex interplay of public health policies, societal reactions, and individual behaviours."
Positive X dataset	"The thematic analysis highlights an ongoing tension between public health policies and individual behaviours. Masks and vaccines are central to the discourse, with significant debate over their adoption and effectiveness. Variants like Omicron continue to influence strategies, while community responsibility emerges as a key element in combating the pandemic. These findings underline the complexity of managing public health in a rapidly evolving situation."
Negative X dataset	"The thematic analysis reveals significant public resistance and scepticism toward COVID-19 measures. Vaccine safety and mask efficacy are hotly debated, often linked with broader concerns about personal freedoms and ethical mandates. Misinformation and a lack of trust in authorities further exacerbate these issues. Public fatigue with ongoing health measures indicates a need for transparent communication and adaptive strategies moving forward."

[a]GA: Google Alerts.

In the positive GA dataset, themes included masks being recommended in particular contexts (eg, "...masks are recommended in health care settings, on public transport, in crowded indoor settings") and encouraging vaccination in at-risk groups (eg, "Aside from the government and GPs, family members have an important role in encouraging senior citizens to get vaccines or booster shots against COVID-19").

In the negative GA dataset, themes included reports of the COVID-19 vaccines causing injuries (eg, "Dr [name] has broken her silence about a 'devastating' COVID vaccine injury, slamming regulators for 'censoring' public discussion, and Thousands of Australians suffering from COVID-19 vaccine injury feel they are 'not being heard' or treated fairly by the government"). Occupational vaccine mandates were also discussed with negative sentiment expressed (eg, "Over 200 firefighters in New South Wales and Victoria are being forced to 'stay away from saving lives' because of ongoing vaccine mandates, and Coles is the only major supermarket in Australia that continues to use discriminatory COVID vaccination mandates for workers").

In the positive X dataset, these included masks being protective for the wearer, with reasons of vulnerability or framing it as "smart" behavior (eg, "My daughter works with COVID patients and we wear a mask when she visits. My partner has cancer, and Lots of people coughing on the train and I'm the only one smart enough to wear a mask"). People also discussed the wearing of masks to protect others in the community (eg, "It is selfish not to wear a mask to protect the health of our most vulnerable"). There were many posts expressing the opinion that the benefits of vaccines outweigh the risks (eg, "I'm sorry for people injured by the vaccine but the risk is insignificant compared to complications from Covid"). When mentioning boosters, the sentiment tended to be favorable, wanting access to updated boosters (eg, "I saw on the news that we might get 5th jab in February – can't come soon enough!").

In the negative X dataset, a strong theme emerged about COVID-19 vaccines causing injuries and deaths. Many tweets included statistics regarding vaccine deaths and anecdotal reports of people they know or had heard of dying suddenly after being vaccinated, or reporting lived experience of having a vaccine injury (eg, "Vaccine injury and deaths outnumber actual deaths FROM Covid," and "My neighbour told me that her son's friend went down to the local shopping center to get vax, dies 15 minutes later"). There was also a theme that COVID-19 vaccines are experimental, and people who take the vaccine are "brainwashed" and "sheep," positioning those who had not had the vaccine as "smart" (eg, "Scientists came up with this vaccine in 10 weeks, and people still believe it's safe, talk about being brainwashed, and The sheep are rolling up their sleeves for their 5th jab, and I don't inject poison into my body, unlike the vax junkies").

XSL•FO
RenderX

Thematic Analysis Results From LLM

The LLM-generated summaries of the thematic analysis of the positive- and negative-assigned GA and X datasets are presented in Tables 7-10. Results of LLM thematic analysis, including theme descriptions and the proficiency and reasonableness scores benchmarked against human-coded thematic analysis, are presented in Table 2. When compared to the human-coded thematic analysis, the results for the LLM proficiency were 13 of 20 proficient and 7 of 20 partially proficient, showing that the LLM produced themes that were relevant to the dataset and closely matched with the human-coded themes. All themes, whether fully or partially proficient at matching human-generated themes, were rated "very reasonable" (the themes had high relevance to the dataset and would likely be generated by human coding), suggesting that the results may be useful to support human coding of large datasets.

XSL•FO
RenderX

Table . Results of the large language model (LLM)–coded thematic analysis of positive and negative Google Alert (GA) results and tweets: themes from the positive GA dataset.

Theme	Description	Example of a representative GA result	Proficiency score[a]	Reasonableness score[b]
COVID-19 vaccination and immunity	• Emphasis on vaccination rollouts, booster campaigns, and their impact on immunity • Subthemes: • Effectiveness and benefits: vaccines reducing ICU[c] admissions and mortality • Booster campaigns: urging booster shots to combat waning immunity • Vaccination messaging: efforts to dispel misinformation and encourage uptake	• "Vaccination against COVID-19 disease aims to reduce the severity of symptoms and need for hospitalizations." • "A booster dose can safeguard against severe illness." • "Vaccination was also found to be a cost-saving measure."	Proficient	2
Mask-wearing policies	• Role of mask mandates in mitigating viral spread • Subthemes: • Mandates and compliance: discussion on mandatory versus voluntary mask use • Context-specific recommendations: masks encouraged in health care and public transport	• "Masks should be compulsory this Christmas to quash a fresh spike in Covid cases." • "Masks are recommended in healthcare settings and crowded indoor spaces."	Proficient	2
Economic and social effects	• Evaluation of public health measures on economic stability and daily life • Subthemes: • Economic savings: prevention of losses via vaccination and lockdown measures • Social changes: adaptations during and after pandemic	• "The vaccine rollout saved the national economy $181bn in potential damage." • "Australia's hospitality venues are recording a boom in revenue as we enjoy post-Covid freedoms."	Proficient	2
Emerging variants and evolving strategies	• Addressing new COVID-19 strains and their implications • Subthemes: • Variant-specific concerns: Omicron and immunity challenges • Hybrid immunity: interplay of natural and vaccine-induced immunity	• "Countries should consider recommending masks on long-haul flights given the spread of Omicron." • "Hybrid immunity has contributed to poor vaccine booster uptake."	Proficient (LLM explores variants and immunity more comprehensively than human coding)	2

XSL•FO
RenderX

Theme	Description	Example of a representative GA result	Proficiency score[a]	Reasonableness score[b]
Public health communication	• Messaging around guidelines and behavioral changes • Subthemes: • Hygiene and ventilation: promoting preventive measures • Community engagement: tailored campaigns for vulnerable populations	• "Good ventilation assists in reducing the risk of spreading airborne diseases." • "Health experts are warning Australians not to miss important screenings and vaccinations."	Proficient (LLM explores messaging around all PHMs[d] more comprehensively than human coding)	2

[a]Proficiency score: proficient, partially proficient, not proficient. and not proficient.

[b]Reasonableness score: 2=very reasonable, 1=reasonable, and 0=not reasonable.

[c]ICU: intensive care unit.

[d]PHM: public health measure.

XSL•FO
RenderX

Table . Results of the large language model (LLM)–coded thematic analysis of positive and negative Google Alert (GA) results and tweets: themes from the negative GA dataset.

Theme	Description	Example of a representative GA result	Proficiency score	Reasonableness score
Vaccine mandates and resistance	• Explores public reactions to vaccine mandates, including compliance and opposition • Subthemes: • Compliance challenges: public debates on mandates for essential workers and specific industries • Resistance and injuries: reports of adverse reactions and public resistance	• "Coles is the only major supermarket in Australia that continues to employ discriminatory Covid vaccination mandates for workers." • "Dr (name) has spoken out about her wife's 'severe' reaction to the COVID jab."	Proficient	2
Mask mandates	• Discussion on the effectiveness, compliance, and psychological effects of mask-wearing • Subthemes: • Effectiveness and compliance: questioning the effectiveness of mask mandates • Public perception: mixed reactions to prolonged mask-wearing requirements	• "More embarrassing for Team Mask, those even more uncomfortable N95 masks made little to no difference either." • "On lone mopeds, driving their own cars, on goes the mask."	Proficient	2
Lockdowns and societal effects	• Highlights the societal consequences of lockdowns, including protests and mental health issues • Subthemes: • Protests and backlash: resistance to prolonged lockdowns • Mental health impacts: isolation and its effects on well-being	• "Fairfield and other areas of western Sydney were subject to strict COVID-19 lockdowns." • "Most young Australians have come away feeling lonely and left out."	Proficient	2
Public health messaging	• Examines the effectiveness of communication around health measures • Subthemes: • Misinformation and clarity: confusion around vaccine safety and mandates • Behavioral recommendations: promoting hygiene and mask use	• "Covid frightbat GP previously railed against vaccine misinformation that scared people away from being vaccinated." • "Masks are no longer suggested in public spaces."	Proficient (LLM explores messaging around PHMs more comprehensively than human coding)	2

XSL•FO
RenderX

Theme	Description	Example of a representative GA result	Proficiency score	Reasonableness score
Emerging variants and vaccination strategies	• Discusses adaptive responses to new COVID-19 variants • Subthemes: • Variant-specific challenges: addressing immunity gaps • Booster campaigns: advocacy for continued vaccination efforts	• "Waning immunity is apparent amongst many communities." • "Booster doses continue to help the most vulnerable even as more contagious variants have popped up."	Proficient (LLM explores variant-specific issues and immunity more comprehensively than human coding)	2

[a]Proficiency score: proficient, partially proficient, not proficient. and not proficient.

[b]Reasonableness score: 2=very reasonable, 1=reasonable, and 0=not reasonable.

XSL•FO
RenderX

Table . Results of the large language model (LLM)–coded thematic analysis of positive and negative Google Alert (GA) results and tweets: themes from positive X dataset.

Theme	Description	Example of a representative tweet	Proficiency score[a]	Reasonableness score[b]
Mask usage and debate	• Conversations about the use of masks as a public health tool • Subthemes: • Mandatory masking: debates over government-mandated mask policies • Efficacy and preferences: discussions on the effectiveness of different types of masks (eg, N95 vs cloth) • Public compliance: observations on varying levels of mask adoption	• "Mandate masks on planes would be the bare minimum." • "N95 masks are your best line of defence. Take care." • "Flew from Newcastle to Brisbane today. Apart from us, only one other family wore masks."	Proficient	2
Vaccination and public perception	• Focuses on vaccination campaigns, uptake, and public attitudes • Subthemes: • Booster campaigns: efforts to encourage additional doses for improved immunity • Safety and side effects: concerns about vaccine side effects and injuries • Resistance and advocacy: voices both supporting and opposing vaccination	• "The public must get the 4th vax and mask up." • "COVID vaccines have caused 14 deaths. These deaths are tragic but were expected." • "Vax works!! Vaccines are definitely a better protection than masks."	Partially proficient (human coding also focused on the benefits of vaccines and effectiveness)	2
Policy and public health strategies	• Explores government policies and public discourse on their effectiveness • Subthemes: • Lockdown measures: reflections on the impact and necessity of lockdowns • Social distancing and hygiene: recommendations for maintaining distance and promoting hygiene	• "We needed a nuanced COVID-19 policy that included indoor masks." • "Masks and social distancing work. It was a good run while it lasted."	Proficient	2

XSL•FO
RenderX

Theme	Description	Example of a representative tweet	Proficiency score[a]	Reasonableness score[b]
Emerging COVID-19 variants	• Discusses the impact of new COVID-19 variants on public health responses • Subthemes: • Omicron and immunity: challenges posed by the Omicron variant • Evolving strategies: adjusting public health measures in response to new variants	• "Waning immunity is apparent amongst many communities." • "Variants mean changing our habits."	Partially proficient (human coding also focused on wanting access to boosters due to new variants, and PHM generally)	2
Community behavior and responsibility	• Focus on individual and collective adherence to health measures • Subthemes: • Risk awareness: acknowledging personal and societal risks • Public responsibility: encouragement for collective action to mitigate risks	• "We wear masks to save others' lives as well." • "Masks are our only weapon against COVID. This method has worked for us so far."	Proficient	2

[a]Proficiency score: proficient, partially proficient, and not proficient.

[b]Reasonableness score: 2=very reasonable, 1=reasonable, and 0=not reasonable.

XSL•FO
RenderX

Table . Results of the large language model (LLM)–coded thematic analysis of positive and negative Google Alert (GA) results and tweets: themes from the negative X dataset.

Theme	Description	Example of a representative tweet	Proficiency score[a]	Reasonableness score[b]
Vaccine safety and side effects	• Concerns about vaccine-induced side effects dominate the conversation • Subthemes: • Personal accounts of side effects: tweets describing adverse events • Skepticism toward safety testing: critiques of vaccine trial processes • Call for accountability: demands for recognition and compensation for injuries	• "My 24-year-old nephew collapsed after the Pfizer vaccine and now suffers from myocarditis." • "Emergency vaccines weren't tested long enough for safety." • "The government must acknowledge vaccine injuries and help those affected."	Proficient	2
Mask effectiveness and resistance	• Heated debate on the effectiveness and necessity of masks • Subthemes: • Efficacy questions: disputes over the scientific basis for masks • Resistance to mandates: opposition to mask-wearing policies • Cultural and behavioral aspects: comments on individual mask-wearing habits	• "Masks don't work to reduce transmission at a population level." • "Stop forcing masks on us. We deserve freedom of choice." • "People wearing masks alone in their cars – what's the point?"	Proficient	2
Economic and social misinformation and distrust	• Mistrust in authorities and allegations of misinformation • Subthemes: • Media and government distrust: criticism of public health messaging • Manipulation allegations: claims of hidden agendas behind health measures • Need for transparency: calls for open sharing of data and studies	• "Mainstream media keeps spreading vaccine propaganda." • "The pandemic was planned to control us through fear and mandates." • "We need real data on vaccine injuries, not censored narratives."	Partially proficient (human coding also focused on the role of pharmaceutical companies, not just media and government)	2

XSL•FO
RenderX

Theme	Description	Example of a representative tweet	Proficiency score[a]	Reasonableness score[b]
Ethical concerns about mandates	• Ethical debates around the imposition of health mandates • Subthemes: • Coercion and freedom: mandates seen as violations of personal liberty • Social divisions: how mandates create tension and polarization • Impact on vulnerable groups: concerns about marginalized communities	• "No one should be forced to take a vaccine against their will." • "Stop dividing us into vaxxed and unvaxxed groups." • "Mandates disproportionately affect those with medical exemptions."	Partially proficient (human coding also focused on the stereotyping of people following government advice as brainwashed)	2
Long-term public health strategies	• Critical evaluation of measures like boosters and lockdowns • Subthemes: • Effectiveness of boosters: mixed reactions to additional doses • Policy fatigue: frustration with prolonged measures • Variant-specific adaptations: adjustments in strategy due to new variants	• "How many boosters are we expected to take? It's exhausting." • "Lockdowns destroyed businesses and mental health." • "Omicron shows we need better vaccines, not just boosters."	Proficient	2

[a]Proficiency score: proficient, partially proficient, not proficient. and not proficient.

[b]Reasonableness score, 2=very reasonable, 1=reasonable, 0=not reasonable.

Discussion

Principal Findings

Our study compared the results of AI analysis of PHM-related datasets with human-coded analysis for common tasks used in the context of health emergencies, such as sentiment analysis (to explore public opinion of PHMs), topic modeling (to identify what is being discussed in online news and social media), and thematic analysis (for a more in-depth analysis of how PHMs are being discussed in the public domain). AI tools were deliberately selected to reflect those accessible to public health professionals in Australia that require minimal technical expertise.

Overall, AI tools performed inconsistently across tasks. All models showed poor performance for sentiment analysis. ChatGPT-4 was found to demonstrate stronger alignment with human raters for both the topic modeling and thematic analysis tasks. These findings highlight both the potential and the limitations of AI tools to complement traditional methods of analysis for public health professionals by providing rapid insights while still requiring human interpretation and oversight [26].

Sentiment Analysis Task

To evaluate the accuracy of accessible AI sentiment analysis tools, a subset of Australian English-language GA and X datasets was analyzed by 2 human raters and 5 AI tools. Sentiment was poorly detected by all AI tools, with no agreement between the human rater and either the GA or X datasets (Table 3). In particular, accuracy was less than 30% for neutral sentiment in the GA dataset and under 50% for positive or negative sentiment in the X dataset. These findings align with previous studies showing that off-the-shelf sentiment analysis tools perform poorly when applied to complex, health-related discourse [54,55].

The human-coded sentiment analysis showed that the GA dataset had a majority of neutral sentiment in discussion about PHMs, while the X dataset was evenly split between positive and negative. These findings highlight differences between the data sources, with the prevalence of neutral sentiment in the GA dataset, suggesting predominantly neutral reporting of current stories and events in online news media. The X dataset showed greater polarity of sentiment, which may indicate that users of X have stronger opinions and are more likely to express them online. Our study demonstrates that in Australia during the study period, mask use and isolation were associated with more positive sentiment, while vaccines, lockdowns, and mandates attracted more negativity. These results indicate that sentiment

XSL•FO
RenderX

analysis may give useful high-level insights regarding public opinion for public health decision-makers while highlighting the need for contextual interpretation by human analysts.

Topic Analysis Task

In our study, topic modeling identified overlapping areas of discussion between datasets, with GA focusing on mask use and vaccine mandates in the context of easing restrictions and X focusing on vaccination campaigns, mask mandates, and isolation requirements. LDA partially aligned with human-coded topics, while the LLM output provided fuller contextualization but was overly generalized. Recent advances in generative AI have shown improved topic matching with human annotators of health-related datasets using LLMs [52]. This may assist in public health responses during health emergencies by improving the understanding of topics that are being discussed in the news and on social media and addressing misunderstandings or concerns with public health messaging [56].

Thematic Analysis Task

Qualitative research methods are promoted as the most suitable approach to gain an understanding of the experiences of individuals during health emergencies, which can be used to inform local public health policy decisions and implementation [10]. Manual interpretation of data, including thematic analyses, presents an extensive time and resource burden [52]. Generative LLMs can analyze and interpret vast amounts of text and have shown good accuracy in generating themes when compared to human analysts, with adequate depth of explanations of themes and inclusion of appropriate quotations, with time savings of several hours or even days [13,52,57]. Some studies have suggested that thematic analysis results generated by LLMs may be best used in collaboration with human coders with domain-specific knowledge [14,58,59].

The human-coded thematic analysis of the GA dataset (Table S8a-d in Multimedia Appendix 1) revealed the ways in which PHMs were being discussed in the news media, including strong support for vaccination of at-risk groups and the use of masks in specific high-transmission contexts, while also reporting on vaccine injuries. Themes that emerged from the X dataset revealed polarized views on the uptake of PHM, with the positioning of compliance with vaccines and mask-wearing as “smart” by supporters of those PHM. From the opposing view, there was much discussion about vaccine harms, including injuries and deaths, and the positioning of compliance with PHMs as a result of “brainwashing.” There were also reports of sudden death from the “experimental vaccine,” which is useful for public health professionals to be aware of when planning campaigns to combat misinformation [60].

In our study, the LLM-generated themes were relevant to the dataset, and the majority were closely matched with the human-coded themes. Even when the themes were only partially matched, they rated high on how reasonable it was for the LLM to generate the theme from the dataset, suggesting that thematic analysis may provide valuable input to inform public health decision-making in a timely way. The summaries generated by the LLM provided an overview of the AI-generated thematic analysis; however, they lacked the specific insights and understanding of the social context of the human-generated summaries.

Error and Bias Analysis

Errors in sentiment classification and theme generation were consistent across tools and datasets in this study, reflecting known limitations of NLP and LLM models when applied to public health discourse [55,61,62]. Misclassification of neutral sentiment was frequent, which may result from narrow sentiment thresholds and training data that are not designed for health-related news and social media [54]. Inability to detect sarcasm is another known limitation of AI tools for sentiment and thematic analysis tasks, when the textual data include positive words to express negative sentiment, highlighting another area where human oversight for contextual nuance is important for accurate analysis [63,64]. While LLMs have demonstrated superior accuracy over other AI tools for tasks such as sentiment analysis, topic modeling, and thematic analysis of public health datasets, they do not match human raters for interpretation and depth of analysis [13,52,55,65]. Fine-tuned or domain-specific LLMs trained on health-related text may improve accuracy, but these models are underrepresented, and substantial annotation and validation efforts will be required to develop appropriately trained models for public health contexts [63-65].

LLM reproducibility presents another challenge, as outputs may vary over time as model parameters and training data are updated by developers. This is known as “model drift” and may complicate longitudinal comparisons and replication of results [66]. Researchers should record model versions and exact prompt wording (as in Table S1 in Multimedia Appendix 1) to enhance transparency and allow future verification.

Manual analysis of online public opinion data may be open to bias through the interpretation of the researcher [67]; however, automated methods lack the capacity to clarify the results of analysis, as may be possible with more traditional methods [66]. While social media can give voice to more marginalized groups, health inequities can be amplified if unrepresentative data are used for analysis [68]. Social bots, which are computer algorithms designed to mimic human interactions on social media, can be used to manipulate public opinion and therefore skew sentiment data [69].

Limitations

There were several limitations of our study. First, regarding the data sources used, GA retrieved the first 2 sentences of the article, which may not give an accurate representation of the sentiment of the article. While the anonymity of data from online social networking sites may have benefits over qualitative or survey data by reducing the impact of social desirability bias [70,71], the awareness of the post being observed by others may make the user less likely to publish unpopular opinions [72]. The anonymity of social media data also impacts the ability to collect demographic information, which can impact how generalizable the results of the analysis [73]. The use of social media data for research is becoming increasingly challenging due to restrictions on access by commercial owners of the platform [74]. The use of untrained NLP sentiment analysis

XSL•FO
RenderX

tools, which were unable to correctly identify both neutral sentiment and sarcasm, was a further limitation of this study. While AI can offer a timely way to provide real-time data, public health professionals may be skeptical of the results without understanding the process and how to interpret the output; therefore, education on its effective use will be necessary for future implementation [75]. Inequities and bias that are present in the training data may be replicated in AI outputs [75]; for example, LLMs trained primarily on content originating from North America and the United Kingdom may misrepresent Australian cultural and linguistic nuances, potentially skewing thematic outputs in subtle ways [75]. As this study was conducted on an English-language dataset, the results of this analysis do not capture sentiment and thematic perspectives from culturally and linguistically diverse communities [1]. This limitation is particularly important in multicultural settings, such as Australia, where attitudes toward PHMs may differ across language groups [1]. For AI techniques to inform public health policies in Australia, they must support an understanding of the actual sentiment in diverse communities. While automatic translation and multilingual models can be used for analysis of text in multiple languages, there is a dearth of available data for analysis in languages other than English in the Australian context.

Conclusions

During health emergencies, there is a need to balance rapid analysis of data with accuracy to support public health decision-making. This study examined the accuracy of 5 AI tools in performing tasks designed to measure public opinion in Australia toward PHMs, such as vaccines, mask mandates, and lockdowns. AI tools were chosen that are widely available across government agencies in Australia. All AI tools were found to perform poorly in a sentiment analysis task of the GA and X datasets when compared to a human rater. AI-generated topic modeling and thematic analysis were conducted using the LLM ChatGPT-4 and compared to human-generated responses. The LLM topic modeling outputs showed a high level of alignment with the human-generated topics, and while understanding of the context of the results was indicated, it was often overly generalized. The LLM output of the thematic analysis task was found to be highly relevant and well matched to the human-generated analysis. Even when the themes were only partially matched to the human-generated themes, the results were classified as reasonable, relevant to the data, and likely to be generated by a human analyst. These findings suggest that AI tools, particularly LLMs, may serve as a rapid triage tool to surface emergent themes from large-scale public datasets, which could then be reviewed or refined by human analysts in time-sensitive policy settings. It is unlikely that AI tools will replace traditional research methods used to investigate attitudes and opinions to epidemic PHMs in Australia; however, there remains an opportunity to use this technology to complement qualitative research techniques used by public health professionals in a cost-effective and timely way in the context of health emergencies.

Funding

This work was supported by National Health and Medical Research Council (grant 2022346 and 1137582) and the Balvi Filantropic Foundation.

Data Availability

Data will not be made available, as per the human research ethics committee submitted protocol and approval.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables.
[DOCX File, 123 KB - ojphi_v18i1e80824_app1.docx]

Multimedia Appendix 2
Code for latent Dirichlet allocation.
[DOCX File, 21 KB - ojphi_v18i1e80824_app2.docx]

References

1. Jacobs LR, Mettler S. Why public opinion changes: the implications for health and health policy. J Health Polit Policy Law 2011 Dec;36(6):917-933. [doi: 10.1215/03616878-1460515] [Medline: 22232417]
2. Tahamtan I, Potnis D, Mohammadi E, Singh V, Miller LE. The mutual influence of the World Health Organization (WHO) and Twitter users during COVID-19: network agenda-setting analysis. J Med Internet Res 2022 Apr 26;24(4):e34321. [doi: 10.2196/34321] [Medline: 35275836]
3. Yu G, Garee M, Ventresca M, Yih Y. How individuals' opinions influence society's resistance to epidemics: an agent-based model approach. BMC Public Health 2024 Mar 20;24(1):863. [doi: 10.1186/s12889-024-18310-6] [Medline: 38509526]

XSL•FO
RenderX

4. Disinformation and public health. World Health Organization. 2024. URL: https://www.who.int/news-room/questions-and-answers/item/disinformation-and-public-health [accessed 2024-02-06]
5. Denniss E, Lindberg R. Social media and the spread of misinformation: infectious and a threat to public health. Health Promot Int 2025 Mar 5;40(2):daaf023. [doi: 10.1093/heapro/daaf023] [Medline: 40159949]
6. Nutbeam D, Milat AJ. Artificial intelligence and public health: prospects, hype and challenges. Public Health Res Pract 2025 Mar;35(1):PU24001. [doi: 10.1071/PU24001] [Medline: 40443074]
7. Alcantara AM, Saeri A. Insights to action: an analysis of the COVID-19 pulse survey. : ANZSOG; 2024 Jul. URL: https://anzsog.edu.au/insights/insights-to-action-an-analysis-of-the-covid-19-pulse-survey [accessed 2025-08-14]
8. Vindrola-Padros C, Chisnall G, Cooper S, et al. Carrying out rapid qualitative research during a pandemic: emerging lessons from COVID-19. Qual Health Res 2020 Dec;30(14):2192-2204. [doi: 10.1177/1049732320951526] [Medline: 32865149]
9. Rapid qualitative research to increase COVID-19 vaccination uptake: a research and intervention tool. World Health Organization. 2022. URL: https://www.who.int/europe/publications/i/item/WHO-EURO-2022-4724-44487-62944 [accessed 2026-03-27]
10. Tremblay S, Castiglione S, Audet LA, Desmarais M, Horace M, Peláez S. Conducting qualitative research to respond to COVID-19 challenges: reflections for the present and beyond. Int J Qual Methods 2021 Jan;20:16094069211009679. [doi: 10.1177/16094069211009679]
11. Taylor B, Henshall C, Kenyon S, Litchfield I, Greenfield S. Can rapid approaches to qualitative analysis deliver timely, valid findings to clinical leaders? A mixed methods study comparing rapid and thematic analysis. BMJ Open 2018 Oct;8(10):e019993. [doi: 10.1136/bmjopen-2017-019993]
12. Dong D, Abramowitz S, Matta GC, et al. A rapid qualitative methods assessment and reporting tool for epidemic response as the outcome of a rapid review and expert consultation. PLOS Glob Public Health 2023;3(10):e0002320. [doi: 10.1371/journal.pgph.0002320] [Medline: 37889886]
13. Prescott MR, Yeager S, Ham L, et al. Comparing the efficacy and efficiency of human and generative AI: qualitative thematic analyses. JMIR AI 2024 Aug 2;3:e54482. [doi: 10.2196/54482] [Medline: 39094113]
14. Lancaster K, Rhodes T, Rosengarten M. Making evidence and policy in public health emergencies: lessons from COVID-19 for adaptive evidence-making and intervention. Evid Policy 2020;16(3):477-490. [doi: 10.1332/174426420X15913559981103]
15. Wang C, Wang X, Wang P, Deng Q, Liu Y, Zhang H. Evaluating public opinions: informing public health policy adaptations in China amid the COVID-19 pandemic. Sci Rep 2024;14(1):5123. [doi: 10.1038/s41598-024-55684-4]
16. Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China. IJERPH 2020;17(8):2788. [doi: 10.3390/ijerph17082788]
17. Hu T, Wang S, Luo W, et al. Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective. J Med Internet Res 2021;23(9):e30854. [doi: 10.2196/30854]
18. Mavragani A, Gkillas K. COVID-19 predictability in the United States using Google Trends time series. Sci Rep 2020 Nov 26;10(1):20693. [doi: 10.1038/s41598-020-77275-9] [Medline: 33244028]
19. Venkatesh U, Gandhi PA. Prediction of COVID-19 outbreaks using Google Trends in India: a retrospective analysis. Healthc Inform Res 2020 Jul;26(3):175-184. [doi: 10.4258/hir.2020.26.3.175] [Medline: 32819035]
20. Leeson W, Resnick A, Alexander D, Rovers J. Natural language processing (NLP) in qualitative public health research: a proof of concept study. Int J Qual Methods 2019 Jan 1;18:1609406919887021. [doi: 10.1177/1609406919887021]
21. Wang Z, Ma Y, Song Y, Huang Y, Liang G, Zhong X. The utilization of natural language processing for analyzing social media data in nursing research: a scoping review. J Nurs Manag 2024;2024(1):2857497. [doi: 10.1155/jonm/2857497] [Medline: 40224767]
22. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. J Med Internet Res 2021 Jun 29;23(6):e24435. [doi: 10.2196/24435] [Medline: 34115608]
23. Yousef M, Dietrich T, Rundle-Thiele S. Actions speak louder than words: sentiment and topic analysis of COVID-19 vaccination on Twitter and vaccine uptake. JMIR Form Res 2022 Sep 15;6(9):e37775. [doi: 10.2196/37775] [Medline: 36007136]
24. White BK, Wilhelm E, Ishizumi A, et al. Informing social media analysis for public health: a cross-sectional survey of professionals. Arch Public Health 2024 Jan 2;82(1):1. [doi: 10.1186/s13690-023-01230-z] [Medline: 38167141]
25. Bragge P, Kellner P, Tsering D, Delafosse V. Use of public opinion data to inform COVID-19 policymaking. ANZSOG. 2024. URL: https://anzsog.edu.au/insights/use-of-public-opinion-data-to-inform-covid-19-policymaking [accessed 2025-09-04]
26. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020;22(4):e16206. [doi: 10.2196/16206]
27. Mallawaarachchi V, Meegahapola L, Madhushanka R, Heshan E, Meedeniya D, Jayarathna S. Change detection and notification of web pages. ACM Comput Surv 2021 Jan 31;53(1):1-35. [doi: 10.1145/3369876]
28. Google Alerts. Google. URL: https://www.google.com/alerts [accessed 2025-03-27]
29. Kelly P. Lessons from the fourth Omicron COVID-19 wave. Advice from the Chief Medical Officer, Professor Paul Kelly [press release]. : Australian Government; 2023 Mar URL: https://www.health.gov.au/sites/default/files/2023-03/lessons-from-the-fourth-omicron-covid-19-wave-chief-medical-officer-professor-paul-kelly.pdf [accessed 2025-07-21]

XSL•FO
RenderX

30. Bour C, Ahne A, Schmitz S, Perchoux C, Dessenne C, Fagherazzi G. The use of social media for health research purposes: scoping review. J Med Internet Res 2021 May 27;23(5):e25736. [doi: 10.2196/25736] [Medline: 34042593]
31. About different types of posts. X Help Center. 2024. URL: https://help.x.com/en/using-x/types-of-posts [accessed 2025-02-28]
32. Microsoft Released Update v2311 (build 17029.20108 c2r) current channel for Microsoft Office 365 Products on 12th December 2023. Microsoft Build 2026. 2023. URL: https://learn.microsoft.com/en-us/answers/questions/5283728/microsoft-released-update-v2311-(build-17029-20108) [accessed 2025-10-24]
33. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]
34. van Atteveldt W, van der Velden M, Boukes M. The validity of sentiment analysis: comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. Commun Methods Meas 2021 Apr 3;15(2):121-140. [doi: 10.1080/19312458.2020.1869198]
35. Joshi KP, Jamadar DC. Statistical software applications and statistical methods used in community medicine and public health research studies. Natl J Community Med 2021;12(03):53-56. [doi: 10.5455/njcm.20210329094615]
36. Bertke SJ, Kelly-Reif K. Introducing LTASR, a new R package based on the NIOSH Life Table Analysis System. Occup Environ Med 2022 Nov;79(11):792-792. [doi: 10.1136/oemed-2022-108462]
37. Heinsberg LW, Koleck TA, Ray M, Weeks DE, Conley YP. Advancing nursing research through interactive data visualization with R Shiny. Biol Res Nurs 2023 Jan;25(1):107-116. [doi: 10.1177/10998004221121109] [Medline: 36029021]
38. Digital Transformation Agency. Australian Government Trial of Microsoft 365 Copilot. Australian Government. 2024 Oct 23. URL: https://www.digital.gov.au/initiatives/copilot-trial/microsoft-365-copilot-evaluation-report-full [accessed 2025-07-11]
39. Azure AI apps and agents. Microsoft. 2023. URL: https://azure.microsoft.com/en-au/solutions/ai [accessed 2025-07-08]
40. Hutto CJ, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. ICWSM ;8(1):216-225. [doi: 10.1609/icwsm.v8i1.14550]
41. Feuerriegel S, Pröllochs N. SentimentAnalysis: dictionary-based sentiment analysis. CRAN Project. 2021. URL: https://cran.r-project.org/web/packages/SentimentAnalysis/index.html [accessed 2025-03-27]
42. OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 technical report. : Publishers City; 2023 URL: https://cdn.openai.com/papers/gpt-4.pdf [accessed 2025-06-28]
43. Roehrick K. CRAN: Package “vader”. : Comprehensive R Archive Network; 2022 URL: https://cran.r-project.org/web/packages/vader/vader.pdf [accessed 2025-03-02]
44. Feuerriegel S, Pröllochs N. CRAN: package “sentimentanalysis”. : Comprehensive R Archive Network; 2023 URL: https://github.com/sfeuerriegel/SentimentAnalysis [accessed 2025-07-12]
45. Tutorial: text analytics with Azure AI services. Microsoft. 2024. URL: https://learn.microsoft.com/en-us/azure/synapse-analytics/machine-learning/tutorial-text-analytics-use-mmlspark [accessed 2025-02-04]
46. ChatGPT (January 2025 version). OpenAI. 2025. URL: https://chat.openai.com [accessed 2025-06-28]
47. Braun V, Clarke V. Thematic Analysis: A Practical Guide: SAGE Publications; 2022. [doi: 10.53841/bpsqmip.2022.1.33.46]
48. R Foundation for Statistical Computing. 2023. URL: https://www.r-project.org/ [accessed 2025-07-12]
49. Wittmann FH. Enhancing thematic analysis with large language models: a comparative study of structured prompting techniques. : Aalto University; 2024 URL: https://www.diva-portal.org/smash/get/diva2:1939104/FULLTEXT02 [accessed 2025-05-19]
50. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
51. Deiner MS, Honcharov V, Li J, Mackey TK, Porco TC, Sarkar U. Large language models can enable inductive thematic analysis of a social media corpus in a single prompt: human validation study. JMIR Infodemiology 2024 Aug 29;4:e59641. [doi: 10.2196/59641] [Medline: 39207842]
52. Allaire JJ. RStudio: integrated development environment for for R. : RStudio; 2020 URL: https://www.r-project.org/conferences/useR-2011/abstracts/180111-allairejj.pdf [accessed 2026-03-27]
53. He L, Omranian S, McRoy S, Zheng K. Using large language models for sentiment analysis of health-related social media data: empirical evaluation and practical tips. medRxiv. p. 2024. [doi: 10.1101/2024.03.19.24304544]
54. Lossio-Ventura JA, Weger R, Lee AY, et al. A comparison of ChatGPT and fine-tuned open pre-trained transformers (OPT) against widely used sentiment analysis tools: sentiment analysis of COVID-19 survey data. JMIR Ment Health 2024 Jan 25;11:e50150. [doi: 10.2196/50150] [Medline: 38271138]
55. Chipidza W, Akbaripourdibazar E, Gwanzura T, Gatto NM. Topic analysis of traditional and social media news coverage of the early COVID-19 pandemic and implications for public health communication. Disaster Med Public Health Prep 2022 Oct;16(5):1881-1888. [doi: 10.1017/dmp.2021.65] [Medline: 33653437]
56. Yang Y, Alba C, Wang C, Wang X, Anderson J, An R. GPT models can perform thematic analysis in public health studies, akin to qualitative researchers. J Soc Comput 2024;5(4):293-312. [doi: 10.23919/JSC.2024.0024]
57. Nguyen-Trung K. ChatGPT in thematic analysis: can AI become a research assistant in qualitative research? Qual Quant 2025 Dec;59(6):4945-4978. [doi: 10.1007/s11135-025-02165-z]
58. Liu A, Sun M. From voices to validity: leveraging large language models (LLMs) for textual analysis of policy stakeholder interviews. AERA Open 2025 Jul;11. [doi: 10.1177/23328584251374595]

XSL•FO
RenderX

59. Muric G, Wu Y, Ferrara E. COVID-19 vaccine hesitancy on social media: building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642. [doi: 10.2196/30642] [Medline: 34653016]
60. Villanueva-Miranda I, Xie Y, Xiao G. Sentiment analysis in public health: a systematic review of the current state, challenges, and future directions. Front Public Health 2025;13:1609749. [doi: 10.3389/fpubh.2025.1609749] [Medline: 40620557]
61. He L, Yin T, Zheng K. They May Not Work! An evaluation of eleven sentiment analysis tools on seven social media datasets. J Biomed Inform 2022 Aug;132:104142. [doi: 10.1016/j.jbi.2022.104142] [Medline: 35835437]
62. Joshi A, Bhattacharyya P, Carman MJ. Automatic sarcasm detection. ACM Comput Surv 2018 Sep 30;50(5):1-22. [doi: 10.1145/3124420]
63. Tan YY, Chow CO, Kanesan J, Chuah JH, Lim Y. Sentiment analysis and sarcasm detection using deep multi-task learning. Wirel Pers Commun 2023;129(3):2213-2237. [doi: 10.1007/s11277-023-10235-4] [Medline: 36987507]
64. Espinosa L, Salathé M. Use of large language models as a scalable approach to understanding public health discourse. PLOS Digit Health 2024 Oct;3(10):e0000631. [doi: 10.1371/journal.pdig.0000631] [Medline: 39401186]
65. Giles EL, Adams JM. Capturing public opinion on public health topics: a comparison of experiences from a systematic review, focus group study, and analysis of online, user-generated content. Front Public Health 2015;3:200. [doi: 10.3389/fpubh.2015.00200] [Medline: 26380248]
66. Zolnoori M, Huang M, Patten CA, et al. Mining news media for understanding public health concerns. J Clin Transl Sci 2019 Oct 23;5(1):e1. [doi: 10.1017/cts.2019.434] [Medline: 33948233]
67. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. BMC Public Health 2022 Nov 22;22(1):2146. [doi: 10.1186/s12889-022-14422-z] [Medline: 36419010]
68. Weng Z, Lin A. Public opinion manipulation on social media: social network analysis of Twitter Bots during the COVID-19 pandemic. Int J Environ Res Public Health 2022 Dec 7;19(24):16376. [doi: 10.3390/ijerph192416376] [Medline: 36554258]
69. Kanchan S, Gaidhane A. Social media role and its impact on public health: a narrative review. Cureus 2023 Jan;15(1):e33737. [doi: 10.7759/cureus.33737] [Medline: 36793805]
70. Bispo Júnior JP. Social desirability bias in qualitative health research. Rev Saude Publica 2022;56:101. [doi: 10.11606/s1518-8787.2022056004164] [Medline: 36515303]
71. Olteanu A, Castillo C, Diaz F, Kıcıman E. Social data: biases, methodological pitfalls, and ethical boundaries. Front Big Data 2019;2:13. [doi: 10.3389/fdata.2019.00013] [Medline: 33693336]
72. Mellon J, Prosser C. Twitter and Facebook are not representative of the general population: political attitudes and demographics of British social media users. Research & Politics 2017 Jul;4(3):2053168017720008. [doi: 10.1177/2053168017720008]
73. Takahashi M, Bettinson M. Analyzing online public discourse in Australia: Australian Twittersphere and NewsTalk corpora. Australian Journal of Linguistics 2024 Jul 2;44(2-3):238-260. [doi: 10.1080/07268602.2024.2380690]
74. Olawade DB, Wada OJ, David-Olawade AC, Kunonga E, Abaire O, Ling J. Using artificial intelligence to improve public health: a narrative review. Front Public Health 2023;11. [doi: 10.3389/fpubh.2023.1196397]
75. Wang W, Jiao W, Huang J, et al. Not all countries celebrate thanksgiving: on the cultural dominance in large language models. 2024 Presented at: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1); Aug 1-16, 2024. [doi: 10.18653/v1/2024.acl-long.345]

Abbreviations

AI: artificial intelligence
API: application programming interface
GA: Google Alerts
LDA: latent Dirichlet allocation
LLM: large language model
NLP: natural language processing
PHM: public health measure
VADER: Valence Aware Dictionary and Sentiment Reasoner

XSL•FO
RenderX

Edited by N Akpakpan; submitted 17.Jul.2025; peer-reviewed by LI Barbosa-Santillan, P Sharma; revised version received 12.Jan.2026; accepted 17.Feb.2026; published 07.Apr.2026.

Please cite as:
Hutchinson D, Lee L, Stone H, Moa A, Seale H, MacIntyre CR
Comparison of Artificial Intelligence Tools With Human Coding for Sentiment, Topic, and Thematic Analysis Tasks of Public Health Datasets During the COVID-19 Pandemic in Australia: Case Study
Online J Public Health Inform 2026;18:e80824
URL: https://ojphi.jmir.org/2026/1/e80824
doi:10.2196/80824

© Danielle Hutchinson, Lauren Lee, Haley Stone, Aye Moa, Holly Seale, C Raina MacIntyre. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 7.Apr.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI-Based Diagnostic Platform Capabilities With Lyme Disease as a Use Case: Integrative Exploration

Sarah P Maxwell[1], PhD; Connie L McNeely[2], PhD; Abdollah Zeraatpisheh[3], PhD; Chris Brooks[4], PhD; Claire Maxwell[5], BS; Kevin C Thomas[6], PhD

[1]Public Health, Public Management, and Public Policy, School of Economic, Political and Policy Sciences, The University of Texas at Dallas, 800 W Cambell, Richardson, TX, United States

[2]Schar School of Policy and Government, George Mason University, Fairfax, VA, United States

[3]Public and Non-Profit Management, School of Economic, Political and Policy Sciences, The University of Texas at Dallas, Richardson, TX, United States

[4]School of Biological Sciences, University of Oklahoma, Norman, OK, United States

[5]O'Donnell School of Public Health, The University of Texas Southwestern Medical Center, Dallas, TX, United States

[6]Laboratory for Human Neurobiology, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States

Corresponding Author:
Sarah P Maxwell, PhD
Public Health, Public Management, and Public Policy, School of Economic, Political and Policy Sciences, The University of Texas at Dallas, 800 W Cambell, Richardson, TX, United States

Abstract

Background: Lyme disease (LD) is the most common vector-borne disease in the United States. It is difficult to diagnose because it can mimic numerous other conditions, and testing protocols may not be sufficient. Although the Centers for Disease Control and Prevention (CDC) recommend a 2-tiered serologic testing approach for LD diagnosis, many patients are diagnosed clinically based on criteria such as the erythema migrans rash and, particularly when the rash is not present, various symptom patterns, exposure history, other information, and clinician observations. Against this backdrop, online symptom checkers, using artificial intelligence (AI) processing techniques, are increasingly used to obtain diagnostic information and resources. With LD as a use case, this research applied a modified capabilities approach to explore the relative effectiveness and utility of AI-based tools in application and comparison to serologically (CDC+) and clinically based diagnoses.

Objective: The overarching goal of this research was to provide a baseline exploration of AI-assisted diagnostic tools relative to more traditional medical assessment approaches in detecting complex infectious diseases. To assess the potential diagnostic utility (DU) of online symptom checker platforms with LD as a use case, this study aimed to (1) evaluate platform performance in identifying LD across different diagnostic cohorts; (2) compare symptom patterns and severity distributions among online LD diagnoses; and (3) identify the most frequently co-occurring conditions potentially misclassified as LD or vice versa.

Methods: Data were drawn from a limited structured survey of patients with confirmed or probable LD, including diagnostic pathways (CDC+ and/or clinical), symptom profiles and severity, treatment history, and time to diagnosis. These patient cases were then entered into 3 leading AI-based symptom checker platforms—MediFind, Isabel, and WebMD—to examine diagnostic performance. Descriptive analytics, logistic regressions, and postestimation analyses were used to identify patterns of DU and interaction effects among cohort type, symptom severity, and AI-based platforms. This survey-based research was not intended to serve as an experiment or clinical trial.

Results: DU varied significantly across platforms and symptom severity thresholds. DU improved at higher symptom severity thresholds (≥3; $P<.001$) and was slightly higher among clinically diagnosed cohorts compared to CDC+ cohorts ($P=.21$). Marginal analyses revealed that clinically diagnosed respondents were more sensitive to changes in severity, but with different levels of platform consistency across conditions. Note that DU was analyzed principally for exploratory purposes.

Conclusions: This research serves as a preliminary and directive step for expanded data collection and a larger, more comprehensive study. The findings suggest that AI-based symptom checkers may supplement early diagnostic reasoning in complex conditions such as LD, particularly when symptom severity is high. Inconsistency across platforms and diagnostic categories highlights the need for algorithmic refinement and standardized validation frameworks to enhance diagnostic reliability in AI-based tools.

(Online J Public Health Inform 2026;18:e87529) doi:10.2196/87529

XSL•FO
RenderX

KEYWORDS

symptom checker; diagnostic accuracy; Lyme disease; serologic testing; clinical diagnosis

Introduction

Background and Research Context

Lyme disease (LD) is the most common vector-borne disease in the United States, with recent estimates indicating increased numbers of new diagnoses and the treatment of approximately 476,000 individuals each year [1]. However, LD presents with a wide range of multisystem symptoms, and patients often face diagnostic difficulties. In fact, given these difficulties, there are some suggestions that the actual number of cases is grossly underestimated [2]. The perniciousness and expanding prevalence of LD—caused primarily by the bacterium *Borrelia burgdorferi* (and sometimes *Borrelia mayonii*) and spread to humans through the bite of infected blacklegged ticks—has been attributed to various factors, including environmental changes leading to wider tick breeding habitats.

At the same time, digitalization, large-scale data collection, and algorithmic engagement have become increasingly integral and expected in the health arena, not to mention across other domains. In particular, society has witnessed increasing developments in digital health resources available to the public (as well as to health professionals) and used as popular means for accessing relevant information, checking symptoms, and determining possible conditions and treatments. Of special note in this regard are symptom-checking applications, that is, online or software tools that provide a platform for patient and clinician use.

Symptom checker platforms allow users to input a variety of symptoms along with physical and other personal characteristics (age, sex, health history, geographic location, etc), which are then analyzed to produce different diagnoses, typically presented as a ranking in terms of diagnostic likelihood and potential seriousness [3-5]. Symptom checkers do not ask users for their medications, vital signs, or other questions that might otherwise be incorporated into a physical examination [4,6,7].

With artificial intelligence (AI) and natural language processing (NLP) as the basic means for analysis, symptom checker platforms typically require a brief medical and demographic history (eg, country of residence, sex, pregnant or not pregnant, and age). AI and NLP algorithms convert patient responses into different diagnoses, often listed in order of potential seriousness. Online symptom checkers use different AI capabilities, offering opportunities for improved diagnostic access and cost efficiency when patients can be diverted from the emergency room (ER) or quickly diagnosed [8,9]. However, performance varies substantially across platforms, reflecting differences in data scope, machine learning models, and linguistic input design.

Diagnostic utility (DU) remains the central issue, and the effectiveness of symptom checker platforms in this regard is a fundamental concern for patients, clinicians, and public health in general [9]. Given their growing adoption, greater reliance on symptom checker platforms for initial guidance and confirming information on LD and other tick-borne diseases can be expected relative to their increasing prevalence. This is particularly important given the variety of possible symptoms and conditions associated with LD and their overlap with other unrelated diseases. Moreover, it points to the issue addressed in this study—the DU of online symptom checker platforms relative to LD as a use case. To address this, a 3-fold interrelated capabilities approach is used: (1) it compares and assesses online symptom checker efficacy in diagnosing LD; (2) it provides a symptomatological comparison relative to online LD diagnoses; and (3) it compares symptoms, other conditions, and information relative to online non-LD diagnoses. Taken together, these comparisons provide for an encompassing analysis of diagnostic concurrence and online symptom checker applications for LD diagnoses.

Research Focus

In today's rapidly evolving digital landscape, the capabilities of online symptom checker platforms also are changing and expanding in response to AI-driven developments. Moreover, online symptom checkers are increasingly accessed—by laypersons, patients, and medical professionals—for diagnostic information and resources. In these days of "big data" and extended algorithmic application and engagement, they are, more than ever, being viewed as authoritative and reliable sources for decision-making and guidance [9]. However, to what extent do these platforms operate as consistent and reliable assessment bases? This issue is explored here using LD in the United States as a use case, especially given the complexities associated with its diagnosis and its growing incidence across the population and country.

After brief discussions of LD testing and of online symptom checker platform characteristics as background, details are provided on the research goals and the data and methods used for this study. The analysis is then presented, along with a discussion of the findings framed relative to questions of DU and AI-based diagnostic platform capabilities. The conclusion provides a summary and suggestions for practical application and further study.

LD Testing and Presentation

LD is difficult to diagnose. A common assumption is that the defining indicator of LD is the tell-tale "bull's-eye" erythema migrans (EM) rash and, especially in endemic areas, the EM rash alone is often used in clinical diagnoses [10]. However, in actuality, many people do not develop the EM rash, and, even when it is present, clinicians may misinterpret or overlook it as indicative of LD. Moreover, testing protocols may not be sufficient. Testing for LD typically relies on either serologic or clinical diagnostic approaches. In particular, the Centers for Disease Control and Prevention (CDC) in the United States recommends a 2-tiered serologic testing with a predetermined number of positive bands indicating LD, referred to as "CDC+." One limitation of this approach is that, following an infection, antibodies can take weeks to generate, and "these tests have limited sensitivity during the early period of infection when people with EM typically seek medical evaluation" [11].

XSL•FO
RenderX

Serologic testing is often negative in early-stage LD. Even those who seek medical attention for the EM rash and receive antibiotics as treatment "may produce overall lower levels of *B. burgdorferi* antibodies, may only produce antibodies against a limited repertoire of *Borrelia* antigens, and may have impeded IgM-to-IgG isotype switching. This may lead to false-negative serologic results during the acute clinical phase when the EM is apparent and occasionally even during the convalescent (post-treatment) phase of illness" [11]. In fact, while CDC guidelines are based on serology, it is common for a patient not to produce antibodies. The resulting test may, therefore, produce a false negative when the test is performed too soon. False negatives may also occur when a patient is immunocompromised, has received early treatment, or has late-stage LD [12,13]. These limitations underscore the diagnostic uncertainty surrounding LD.

All in all, inconsistent symptom presentation and testing for LD are a particular concern, given the false negative and false positive occurrences. A lack of antibody production during testing can be problematic, leading some patients to seek out specialty laboratories for clinical testing. At least 3.4 million LD tests are performed by specialty laboratories in the United States each year [14,15]. Still, testing itself remains an issue of debate, especially since the Lyme bacterium is difficult to culture [16]. Moreover, LD is often called "the great imitator," given its symptom similarities with other diseases and its multisystem nature, often affecting multiple organ and tissue systems. It is not unusual for such circumstances to lead to decision fatigue for clinicians [17]. Note, too, that misdiagnoses are more likely in urgent care and ER settings and tend to occur due to symptom misattribution, intermittent symptoms, and the belief of some physicians that an EM rash must be present [18]. Additionally, clinical LD diagnoses sometimes lead to further controversy, as clinicians may fail to test for related co-infections and other diseases with similar presentations [19]. Collectively, these challenges highlight why symptom-based and algorithmic approaches are increasingly considered in LD evaluation.

A final point is that patients diagnosed with LD often develop chronic or long-term symptoms. Lyme infection–associated chronic illnesses can occur after acquiring LD, often presenting as "debilitating physical symptoms including chronic fatigue, recurring pain, cognitive dysfunction such as 'brain fog,' and sleep disturbances despite antibiotic treatment for Lyme disease. Approximately 10 - 20 percent of the 476,000 individuals who develop Lyme disease each year…go on to develop Lyme-IACI" [11]. Persistent postinfection symptoms pose significant challenges for patients and clinicians, underscoring the need for early and accurate diagnosis. Patients with LD often report a lower quality of life and greater health care usage than those with other chronic illnesses [20-22]. This study offers previously untested variables, using both serological and clinical diagnoses among patients with varying degrees of symptom severity prior to COVID-19.

AI Symptom Checker Platforms

Symptom checker online platforms apply NLP, a subfield of AI, to analyze symptoms along with other possibly applicable information entered by the user. Platforms require typical background information such as a basic history of diagnoses, country of residence, and other demographic details, and NLP techniques and AI algorithms are applied to patient responses to determine possible diagnoses and health impacts. Regarding their general use, assessments of symptom checker accuracy against the final diagnosis by various medical providers—ranging from emergency department personnel to general and specialized residents and surgeons—have been highly variable [9]. Platforms also vary in their search features and analytical approaches. For example, studies have shown that some platforms use only 1 symptom (which can be expressed at various levels of detail, from general to fine-grained, and use various expressive descriptors and delineations), whereas others use as many as 20 for analysis [9].

Moreover, depending on the particular platforms and the medical conditions in question, the highest efficacy rate remains relatively low, but symptom checker platforms are continuously being updated and are advancing rapidly via AI and machine learning developments. Technological advances continue to improve algorithmic reasoning and language comprehension, although standardization across platforms remains limited. Currently, reported diagnostic accuracy rates for symptom checkers range from approximately 3% to 53%, depending on the condition set and model design [9].

However, symptom checkers have been used as useful tools in prescreening for diseases such as COVID-19 [23]. In a recent study of diagnostic accuracy of symptom checkers for COVID-19, only 2 of the 10 platforms balanced sensitivity (true positives) and specificity (patients without the disease), while most of the other checkers were particularly sensitive. Overly sensitive symptom checkers resulted in the classification of almost all patients as COVID-19–positive [23]. Other studies use comparisons with physician diagnoses rather than laboratory tests, noting that, in an emergency department setting, "the overall performance of Ada on its top 5 diagnoses compared with the ED diagnoses was not significantly different from that of the study physicians assessed on their top 3 diagnoses" [4]. Other studies indicate that diseases with atypical presentations or those that are less common tend to have lower diagnostic accuracy [24].

The NLP programs themselves can also have important effects on how the platforms are used and on the results they provide. NLP models have demonstrated capabilities for understanding and responding to human input [25], but outcomes can vary depending on particular programs and applications. For example, one widely used platform called Isabel uses more everyday or layperson "common" language [7], whereas others, such as the well-known WebMD, use more medically focused language [6,7,26-29]. Thus, the term "brain fog" does not appear as a symptom in WebMD; users would need to enter an alternative such as "altered mental state," which may or may not occur to them or seem like an accurate description from their perspective. NLP has been applied to analyzing symptoms, such as depression, with a range of recognized conditions and outcomes. Accordingly, in this case, based on NLP, generative AI programs (eg, ChatGPT) have been used in determining and supporting

XSL•FO
RenderX

the appropriate standard of care for patients with both mild and severe depression [30]. These applications illustrate both the promise and limitations of AI-assisted health technologies in translating subjective experiences into diagnostically meaningful data.

Methods

Study Design and Data Sources

This study draws on the extensive review conducted by Wallace et al [9] on the diagnostic and triage accuracy of digital and online symptom checker tools as a foundation for examining their utility in relation to LD. Diagnostic determinants in this regard are based primarily on symptoms [29], patient entry point (eg, ER vs non-ER) [7], or physician specialty [27]. This research also incorporates multiple symptom checker applications with more common-language and medical-language programs, which may vary across platforms. Focusing on patients with LD, it more specifically compares individuals with serologically diagnosed (CDC+) versus clinically diagnosed LD based on tests performed in commercial laboratories.

Individual patient data were drawn from an online survey conducted in 2019 (prior to the onset of the COVID-19 pandemic), disseminated via Facebook and shared among various LD-affiliated organizations and individuals. This resulted in a convenience sample of 67 participants. We recognize that this recruitment strategy, which relied on capturing patients from various LD groups, can introduce selection bias into the sample. For example, it is conceivable that this approach favors patients with chronic, atypical, or polysymptomatic presentations. However, as an exploratory, initializing, and directive step, this approach still offers a fundamental contribution to the literature, as previous symptom checker studies tended to focus on definitive diagnoses rather than conditions that are difficult to diagnose.

Patients who self-reported an LD diagnosis were asked to document their diagnostic type (CDC+ and/or clinical). CDC+ was defined for the respondents using the National Notifiable Diseases Surveillance System [31]. Patients were asked to report their symptoms and symptom severity, as well as other diagnostic-related and treatment-related information, including the time taken to receive a diagnosis and the related financial costs. As shown in Textbox 1, a wide range of LD symptoms has been identified by the CDC and in the medical literature [32], accounting for much of the associated diagnostic difficulty. Additionally, some language or terms used by patients (eg, “brain fog”) do not appear explicitly as CDC-recognized symptoms. This research endeavored to include such symptoms as reported by patients for a more comprehensive symptom delineation and recognition.

XSL•FO
RenderX

Textbox 1. Identified Lyme disease symptoms.

Centers for Disease Control and Prevention (CDC)–recognized symptoms

- Rash
- Fever or chills
- Fatigue
- Malaise
- Headache
- Myalgia
- Arthralgia
- Pharyngeal erythema
- Lymphadenopathy
- Baker cyst
- Conduction abnormalities
- Myocarditis or pericarditis
- Bell palsy or other cranial neuropathy
- Meningitis
- Motor and sensory radiculoneuropathy and mononeuritis multiplex
- Subtle cognitive difficulties
- Encephalitis, encephalomyelitis, subtle encephalopathy, and pseudotumor cerebri

Additional symptoms from the medical literature

- Confusion or altered mental status
- Pain
- Seizures
- Vertigo or dizziness (not reported by CDC)
- Tingling or numbness (not reported by CDC)
- Impaired cognitive function (concentration, memory difficulty, and word recall; not reported by CDC)
- Paralysis: difficulty swallowing (dysphagia) or Bell palsy
- Difficulty with, or slurred speech (dysarthria; not reported by CDC)
- Fainting (syncope; not reported by CDC)
- Depression
- Anxiety (not reported by CDC)
- Mania, panic attacks, delusions, or hallucinations (not reported by CDC)

During July and August 2023, patient-reported symptoms and other relevant information were entered into 3 online symptom checker platforms—MediFind, Isabel, and WebMD. These platforms were selected based on their high levels of recognition, frequency of use, variability in approach, and reported utility [33]. Respondent symptoms were entered according to severity and diagnostic type (clinical or CDC+). MediFind is an online medical website providing medical information and a symptom checker tool designed to identify diseases from peer-reviewed journals, clinical trials, and medical presentations. It is based on NLP and a wide-scoping AI algorithm formation. The MediFind website indicates that its proprietary algorithms use the simplified terminology, allowing users to enter symptoms in “natural” or common language, as well as more medically grounded language [34]. The Isabel symptom checker website states that it is “recognized worldwide for its utility covering both common and rare conditions” [33], suggesting potential in relation to difficult-to-diagnose diseases such as LD.

In 2018, WebMD launched upgraded technology that allows users to enter “easy-to-understand” language into the platform, although levels of applicability and understanding could vary substantially by user. However, compared with other platforms, WebMD was more limited in language and expression and did not offer common-language options for many symptoms. The following changes and observations were made while entering symptoms into WebMD, given the more technical medical language restrictions of the program. These substitutions were

XSL•FO
RenderX

applied consistently to ensure cross-platform comparability of input terms. All symptoms were those listed by the respondents. The change to medically indicated language (eg, "brain fog" to "altered state of mind") was entered based on the occurrence of the only comparable symptoms from which to select such that the medical language was standardized via the online platform.

- "Intermittent pain" and "shooting pains," as described by patients, were not recognized by WebMD. These symptoms were replaced with "pain."
- "GI or stomach problems" was replaced with "nausea."
- "Brain fog" was replaced with "altered state of mind."
- WebMD did not have an option for "panic attacks." However, "anxiety" in general was an included symptom.
- "Extreme fatigue" was substituted with "fatigue."

As mentioned earlier, data entry for this study took place in 2023. At that time, none of the symptom checkers allowed input on symptom severity. Thus, symptom severity was split for each individual respondent into ≤2, which was a natural distinction in minor or more severe symptom presentation, and ≥3; these are actually nested in that ≥3 includes the initial ≤2. Symptom severity for each symptom was entered as follows: first, all symptoms with a severity score of ≥3 were entered, and then, keeping the most severe symptoms in the symptom checker program, the remaining reported symptoms were added for each respondent with a severity ≤2 to the original list. Two sets of diagnoses were produced by each symptom checker platform: one for the most severe symptoms only, and a second set of diagnoses for a respondent's full list of symptoms. Finally, the top 10 diagnoses were obtained by severity and by clinical or CDC+ diagnoses. (If "COVID" resulted as a symptom checker output diagnosis, it was removed.)

In general, analyses were conducted on a "Lyme Disease Artificial Intelligence Diagnosis" dataset constructed for this study, as described above. Constructing the raw LD and AI dataset included unifying the 3 subdatasets (one for each AI platform), cleaning and standardizing the data, and encoding it numerically to facilitate analysis. This dataset consists of self-reported symptoms and AI-generated diagnoses, with the main analysis focused on characterizing potential differences in DU across 3 major categorical variables—cohort, symptom severity threshold, and AI platform. These variables are defined in the following sections.

Dependent Variable

The dependent variable examined in this study was DU.

- DU is determined by the proportion of principal diagnoses that include LD. It is dichotomous, indicating values of "No" and "Yes." It is the proportion of these values across defined groups.

Independent Variables

Three independent variables were included in the analyses.

- *Cohort* defines which LD diagnostic cohort was input into the diagnostic AI platform. It has 2 values: "CDC+" (those diagnosed using CDC serology testing) and "Clinical" (those diagnosed using clinical testing).
- *Symptom severity threshold* (SSTCat) is based on each threshold determined by the number of symptoms as entered into the diagnostic AI platforms. It has 2 values: "Less than or equal to 2" (LT2,≤2) and "Greater than or equal to 3" (GT3,≥3).
- *AI platform* defines which AI-based symptom checker provided the diagnoses. It has 3 values: "MediFind" (MF), "Isabel" (IS), and "WebMD" (MD).

In total, there are 2 × 2 × 3=12 subcategories definable by these categorical variables (CatVars), each with its own set of DU data.

Analysis was undertaken to determine the potential effects that may influence the ability of an AI-based symptom checker to accurately diagnose an individual with LD. In addition to descriptive analytics, logistic regressions were performed to assess the association among variables and to classify and predict associated outcomes. Moreover, postestimation analysis was used for follow-up validity checks on platform comparisons and to identify the principal effects. Analyses were aimed at (1) assessing and comparing the DU of the 3 online diagnostic tools, (2) evaluating diagnostic concurrence with symptoms and test type relative to AI-based outcomes, and (3) comparing variance in LD as opposed to non-LD diagnoses given matching symptom and other information input across platforms. Collectively, these analyses were used to evaluate both platform-specific and cross-cohort diagnostic consistency.

Ethical Considerations

Institutional review board approval (IRB-21-149) for the study protocol was obtained prior to the survey under the formal adoption of the Declaration of Helsinki. All responses were completely anonymous, and survey respondents consented to voluntarily participate in the survey. All survey respondents provided informed consent. No personally associated identification data were collected or reported as such. Participants were not compensated. Data are available upon request.

Results

All sample respondents had an LD diagnosis. A total of 65 participants constituted the initial study sample, of which 39 were in the CDC+ diagnosis cohort and 26 were in the clinical diagnosis cohort. The study sample consisted of 58 (89.2%) female participants and 7 (10.8%) male participants, with an average age of 41.8 (SD 15.3) years (range: 8 - 73). The CDC+ cohort consisted of 34 (87.2%) females and 5 (12.8%) males, with an average age of 40.4 (SD 16.5) years (range: 8 - 73). The clinical cohort consisted of 24 (92.3%) females and 2 (7.7%) males, with an average age of 44 (SD 13.4) years (range: 19 - 72).

Given this sample, a comparative baseline for evaluating AI-driven diagnostic performance in LD across diagnostic categories was derived relative to the research framework, as shown in Figure 1, which summarizes the diagnostic distributions and categorical variable structure forming the basis for the overall analysis. Moreover, these symptom patterns

XSL•FO
RenderX

identified in the survey informed the baseline comparison of diagnostic outputs, as visually summarized in Figure 2.

Figure 1. Diagnostic utility relationships among cohort type, symptom severity thresholds, and artificial intelligence platform.

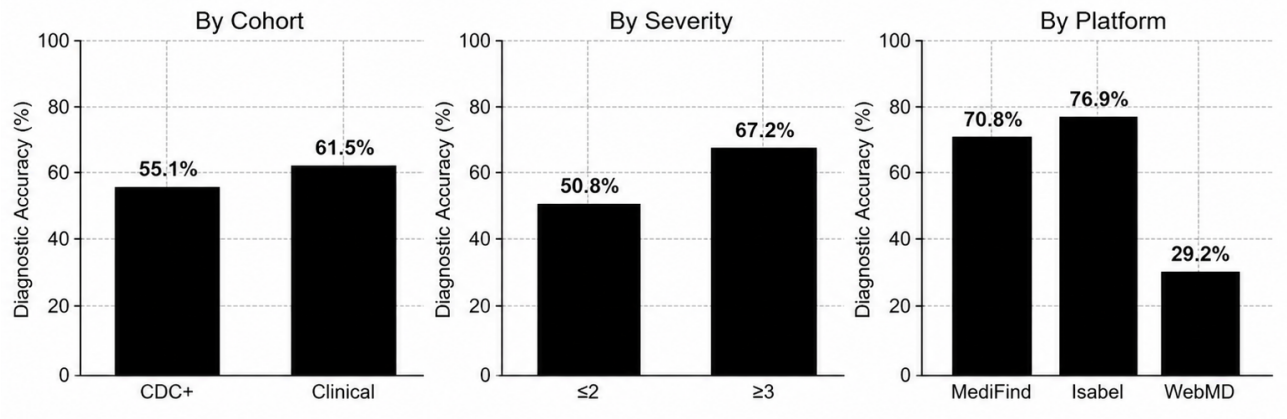
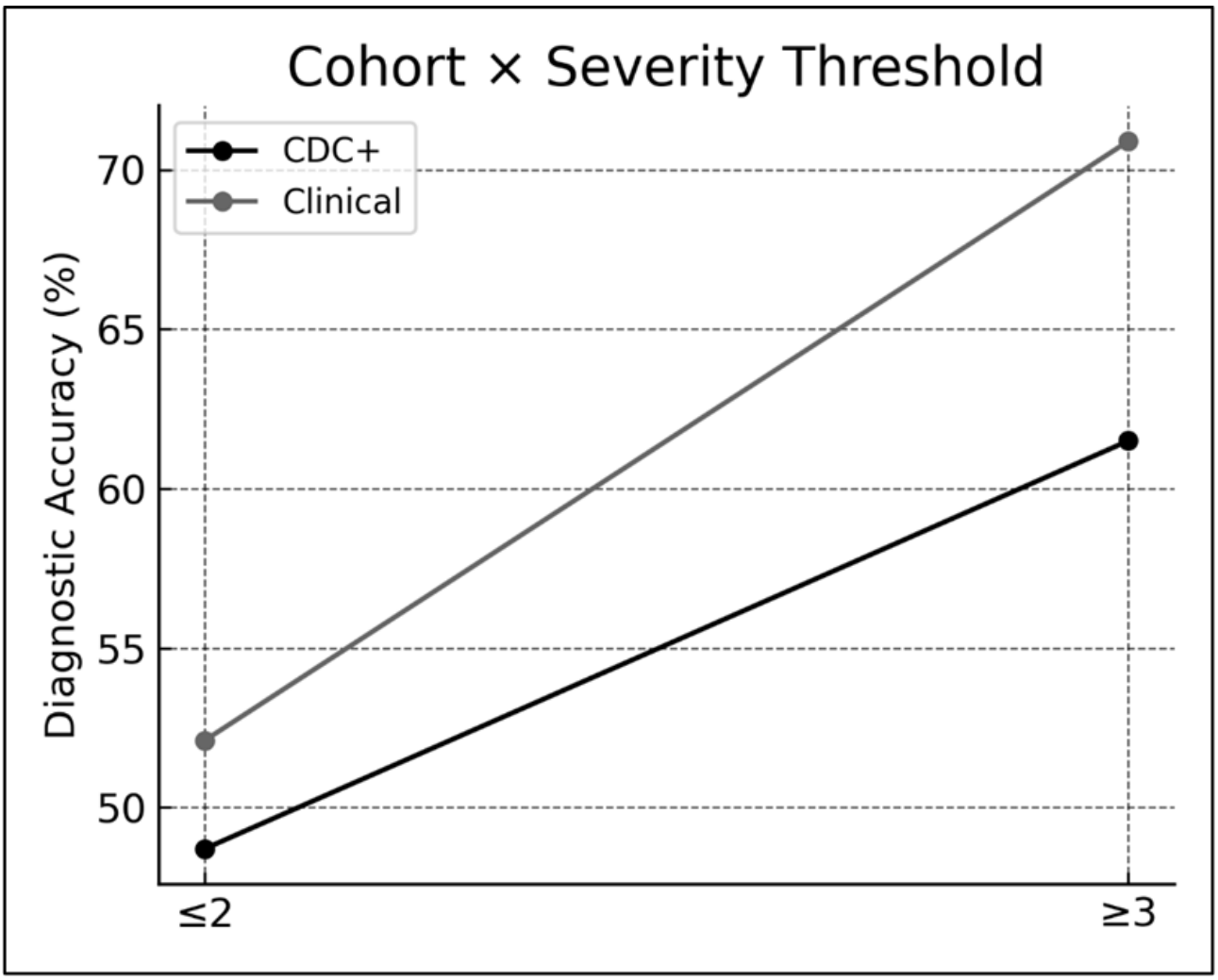


Figure 2. Frequency of Lyme disease diagnoses by cohort and symptom severity.



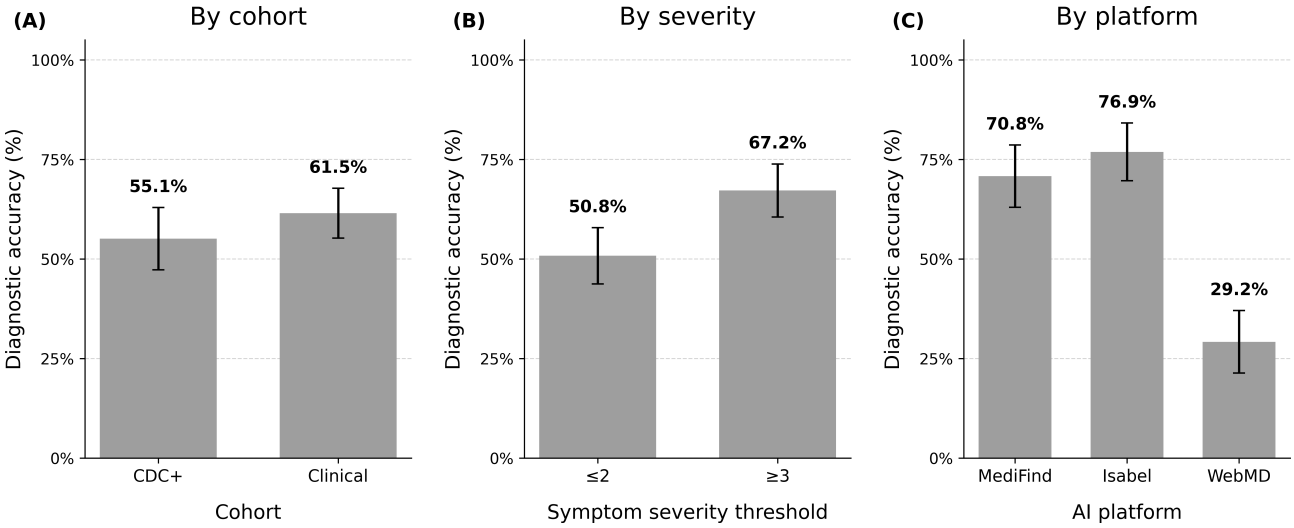
Given the research aims for this study, the following questions were posed to guide the analysis: (1) "Does DU significantly vary within the 3 categorical variables (CatVars)?" and (2) "Are there any interactions between categorical variables (CatVars) that significantly influence DU?" (Null hypothesis: DU does not significantly vary within or between the CatVars). The main objectives were to compare and assess similarities and differences between those with clinical versus CDC+ diagnoses and to determine the potential capabilities and efficacy of online symptom checkers relative to their potential for and levels of DU. To address these, 6 logistic regression models were fit: 3 univariate models for each individual categorical variable and 3 bivariate models for each possible pairing of the categorical variables. Odds ratios, z values, and P values, marginal estimates, and pairwise comparisons were extracted from each

XSL•FO
RenderX

model (see Table S1 in Multimedia Appendix 1 for complete listing of model outcomes and comparisons).

The univariate models each analyzed differences within a single categorical variable, without accounting for the other 2 categorical variables. Figure 3 illustrates the marginal DU estimates (with 95% CIs) by cohort, symptom severity threshold, and AI platform. DU did not vary significantly by cohort (*P*=.21), was significantly higher when the symptom severity threshold was greater than 3 (*P*<.001), and was significantly lower in WebMD (*P*<.001) relative to the other symptom checkers. MediFind and Isabel were not significantly different from each other (*P*=.26). Differences in DU could be due to not just database access, but also the design of the AI algorithms used and NLP learning effects, which may have implications for the capabilities of any AI-based symptom checker.

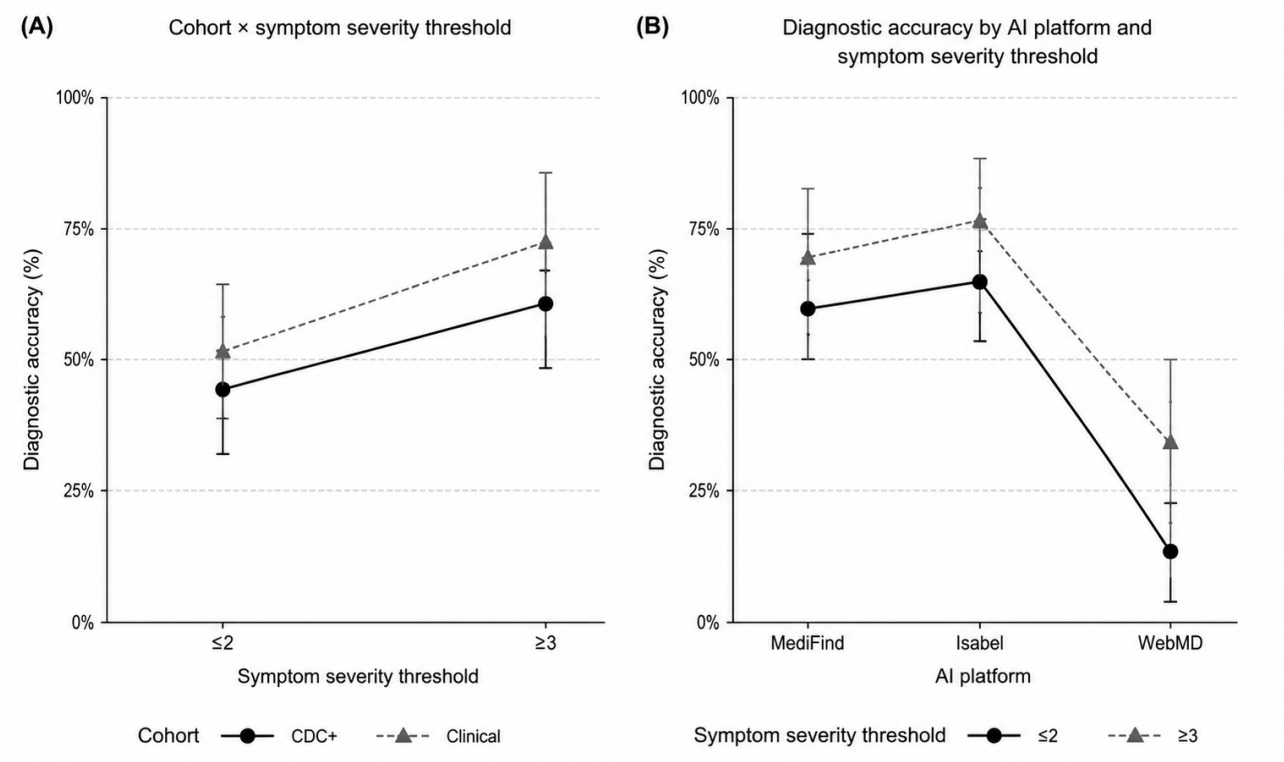
Figure 3. Univariate model results—diagnostic utility by artificial intelligence (AI) platform, cohort, and symptom severity threshold.



The bivariate models analyzed interactions between given pairs of categorical variables and included the main effects for each variable as well as the interaction effect between them. No significant interaction effects were observed in the base models (*P*>.12). Marginal estimates—predictions of the true value of DU in different combinations of the categorical variables—also were generated and evaluated. Several potential interaction effects were observed and subsequent pairwise comparisons were conducted to determine the magnitude and significance of these differences. Postestimation analyses also found a number of significant interactions. Figure 4 shows the interaction effects with 95% CIs, where Figure 4A shows cohort by symptom severity threshold, and Figure 4B shows symptom severity threshold by AI platform. Overall, regarding interaction effects across the AI platforms and cohorts, significantly higher DU was found with the ≥3 threshold than the ≤2 symptom severity threshold, but only in the clinically diagnosed LD cohort (*P*<.001). While there was no significant difference between the ≤2 and ≥3 symptom severity thresholds within the CDC+ cohort (*P*=.12), the difference became significant within the clinical cohort (*P*=.003), suggesting that the clinically diagnosed cohort may be more sensitive to changes in symptom severity. More specifically, the clinically diagnosed cohort received significantly higher DU than the CDC+ group, but only on the Isabel platform (*P*=.09). Isabel and MediFind had significantly higher DU, but only with the ≥3 symptom severity threshold (*P*[IS]=.04; *P*[MD]=.008). It may be that the different trends between the test platforms are canceling each other out, obfuscating interaction effects between diagnostic cohort and the AI applications, a point that merits further investigation and specification relative to AI-based diagnostic capabilities.

XSL•FO
RenderX

Figure 4. Bivariate model results—interaction effects between cohort type and symptom severity threshold: (A) cohort by symptom severity threshold and (B) symptom severity threshold by AI platform. AI: artificial intelligence; CDC: Centers for Disease Control and Prevention.



Note that the significant differences between the ≤2 and ≥3 symptom severity thresholds observed in the univariate models were nonsignificant in the bivariate models (*P*>.11). The largest and most consistent difference in DU was observed between the WebMD and the MediFind or Isabel platforms. While MediFind and Isabel had similar DUs, that of WebMD was significantly lower. Although the ≥3 symptom severity threshold may have significantly higher effects than ≤2, especially when restricted to the clinically diagnosed LD cohort, some of the difference appears to be due to interactions with the particular AI-based platform and cohort, as indicated by the ≤2 and ≥3 interaction main effect becoming nonsignificant in the bivariate models. The clinically diagnosed LD group was notionally higher than the CDC+ group, but this only became significant when restricted to the Isabel AI data. In general, both the MediFind and Isabel online symptom checkers varied significantly between the ≤2 and ≥3 symptom severity thresholds, suggesting an interaction. While there is some level of interaction between AI platforms and symptom severity threshold, greater DU is found with ≥3 than ≤2; however, MediFind appears to benefit in this case less than Isabel or WebMD. This may be an artifact due to MediFind having the highest baseline DU among them, suggesting the need for additional research controlling for such points.

Discussion

Principal Findings

The main objective of this study was to explore AI-assisted diagnostic tools relative to more traditional medical assessment approaches in detecting complex infectious diseases. The focus was on the potential DU of online symptom checker platforms with LD as a use case. A separate goal was to explore nuanced difficulties and contrasts with AI symptom checker input (eg, terminological differences and effects), which shed light on the lack of standardization among various AI platforms. The functionality of the AI platforms introduced unexpected complexities to the study, indicating the need for further research and exploration, particularly as AI develops and its use in medical settings expands.

However, even given the inherent limitations of the platforms, the potential for DU and detection is encouraging. Based on the logistic regressions, along with postestimation, testing for robustness and considering relative influences on the univariate and bivariate models, the major outcomes included a significantly lower DU on the WebMD platform—which could be due principally to differences in AI development and application—and a potentially higher DU in the ≥3 symptom severity threshold. No significant main effects were observed between the cohorts. Interactions among the categorical variables were at times significant, although it is unclear if they were spurious or evidence of a weak effect. This is something that should be further explored in future analyses. In any case, together, the findings underscore the value of comparing AI-based symptom checker performance through a multivariate framework and inform the broader discussion of AI diagnostic reliability in complex conditions such as LD.

The findings here highlight the point that, while DU varies across AI symptom checkers, these tools show measurable promise in differentiating cases of LD under specific symptom severity conditions. AI-based platforms may have the potential to aid in improved diagnostics for conditions and diseases such as LD for which diagnosis is notably challenging. Different from most studies that focus principally on results from clinician

XSL•FO
RenderX

differential diagnosis or imaging, this research also incorporates reported serological findings for a more comprehensive perspective and analysis. To that end, a goal of this research was to determine and examine the fundamental associations between the categorical groups and the primary outcomes in order to inform future analysis and the role of, and reliance on, AI diagnostic platforms. Along those lines, this study contributes to the AI-diagnostic knowledge base and also suggests promise for individuals with lingering symptoms and no definitive diagnosis, as is often the case with LD.

Limitations

Further research is needed with an eye toward conducting a more methodologically exacting assessment of the relevant issues. While this study serves as a preliminary directive exploration focused on the sensitivity of symptom checkers, it did not include a control group for comparison with participants without an LD diagnosis. Both sensitivity, which is the true-positive rate, and specificity, which can identify true negatives, can be problematic in survey designs, especially given the reliance on patient self-reported data. As mentioned, the sample recruitment approach also presented potential limitations, as it may favor patients with chronic, atypical, or polysymptomatic presentations. Note that regarding actual diagnoses and data quality, the survey responses might reflect a certain degree of recall bias, particularly in relation to symptom severity reported for past events, potentially influencing the correlation between reported severity and DU. In addition, the internal validity of the study could be compromised since the definitions used for clinical diagnoses may conflict with or undermine differing AI platform informational approaches relative to potentially unconfirmed diagnoses. Moreover, this points to AI developments and changing capabilities as a critical consideration affecting the potential feasibility and utility of symptom checker platforms over time. However, as noted, this study was exploratory and directive in purpose.

Conclusions

AI diagnostic outputs depend not only on input structure but also on algorithmic design and NLP interpretation, indicating that performance gaps among platforms may reflect both data and model architecture. Attending to these kinds of issues is essential for refining AI-driven diagnostic frameworks and improving their reliability in relation to complex, multisystemic conditions such as LD. As indicated in the analysis, AI platform and symptom severity threshold appear to be the most influential of the categorical variables. Accordingly, future studies should generally account for them in the models that they develop. It may also be appropriate to account for diagnostic type and cohort in some models, as was done here. Moreover, although not pursued in this initial analysis, consideration of secondary variables, as mentioned, may provide a finer-grained understanding of related issues. Categorizing symptoms and diagnoses into several broad subgroups could allow for more granular analyses. Doing so would be trivial for symptoms that have predetermined groups (eg, physical, neurological), whereas diagnoses would require defining a categorization system and then applying it uniformly across different AI platforms, each of which could have its own list of unique diagnoses. In addition to data verification and the use of demographic data to account for intersubjective variability, directly incorporating details on symptoms and non-LD diagnoses can introduce and help to address further delineating questions. For example, what specific symptoms are most associated with successfully reaching an LD diagnosis (or not)? Especially in light of its mimicry character, which diagnoses are most often made in lieu of LD (ie, for which conditions is LD typically mistaken)? Which diagnoses are most often made alongside LD (ie, with which conditions is LD most often associated)? Such questions can be more directly addressed in light of different diagnostic capabilities associated with AI sources and algorithmic developments.

Acknowledgments

Artificial intelligence was not used to write any portion of this paper.

Funding

The authors declared no financial support was received for this work.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: SPM, CLM, KCT
Data curation: CM, CB, AZ, SPM, CLM
Formal analysis: CB, AZ
Investigation: SPM, CLM, KCT
Methodology: AZ, CB, SPM, CLM, CM
Project administration: SPM, CLM
Resources: SPM, CLM
Supervision: SPM, CLM, KCT

XSL•FO
RenderX

Validation: SPM, AZ, CB, CLM, KCT
Visualization: AZ, CLM
Writing – original draft: SPM, CLM, AZ
Writing – review & editing: SPM, CLM, AZ, CM, KCT
All authors have read and agreed to the published version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary analyses of marginal estimates and main and interaction effects.
[DOCX File, 18 KB - ojphi_v18i1e87529_app1.docx]

References

1. CDC provides estimate of Americans diagnosed with Lyme disease each year. Centers for Disease Control and Prevention (CDC). URL: https://archive.cdc.gov/www_cdc_gov/media/releases/2013/p0819-lyme-disease.html [accessed 2026-06-23]
2. Bradshaw MJ, Bloch KC. Tick-borne infections of the central nervous system. In: Hasbun R, Bloch K, Bhimraj A, editors. Neurological Complications of Infectious Diseases: Springer Publications; 2021:325-349. [doi: 10.1007/978-3-030-56084-3_16]
3. Ćirković A. Evaluation of four artificial intelligence-assisted self-diagnosis apps on three diagnoses: two-year follow-up study. J Med Internet Res 2020 Dec 4;22(12):e18097. [doi: 10.2196/18097] [Medline: 33275113]
4. Fraser HSF, Cohan G, Koehler C, et al. Evaluation of diagnostic and triage accuracy and usability of a symptom checker in an emergency department: observational study. JMIR mHealth uHealth 2022 Sep 19;10(9):e38364. [doi: 10.2196/38364] [Medline: 36121688]
5. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021;16(7):e0254088. [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
6. Gilbert S, Mehl A, Baluch A, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269. [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
7. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 8;351:h3480. [doi: 10.1136/bmj.h3480] [Medline: 26157077]
8. Andrew A. Potential applications and implications of large language models in primary care. Fam Med Community Health 2024 Jan 30;12(Suppl 1):e002602. [doi: 10.1136/fmch-2023-002602] [Medline: 38290759]
9. Wallace W, Chan C, Chidambaram S, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med 2022 Aug 17;5(1):118. [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]
10. Nelson C, Elmendorf S, Mead P. Neoplasms misdiagnosed as “chronic Lyme disease”. JAMA Intern Med 2015 Jan;175(1):132-133. [doi: 10.1001/jamainternmed.2014.5426] [Medline: 25365479]
11. National Academies of Sciences, Engineering, and Medicine. Charting a Path Toward New Treatments for Lyme Infection-Associated Chronic Illnesses: The National Academies Press; 2025.
12. Ismail N, McBride JW. Tick-borne emerging infections: ehrlichiosis and anaplasmosis. Clin Lab Med 2017 Jun;37(2):317-340. [doi: 10.1016/j.cll.2017.01.006] [Medline: 28457353]
13. Madison-Antenucci S, Kramer LD, Gebhardt LL, Kauffman E. Emerging tick-borne diseases. Clin Microbiol Rev 2020 Mar 18;33(2):e00083-18. [doi: 10.1128/CMR.00083-18] [Medline: 31896541]
14. Conant JL, Powers J, Sharp G, Mead PS, Nelson CA. Lyme disease testing in a high-incidence state: clinician knowledge and patterns. Am J Clin Pathol 2018 Feb 17;149(3):234-240. [doi: 10.1093/ajcp/aqx153] [Medline: 29425271]
15. Pritt BS, Respicio-Kingry LB, Sloan LM, et al. Borrelia mayonii sp. nov., a member of the Borrelia burgdorferi sensu lato complex, detected in patients and ticks in the upper midwestern United States. Int J Syst Evol Microbiol 2016 Nov;66(11):4878-4880. [doi: 10.1099/ijsem.0.001445] [Medline: 27558626]
16. Rayment C, O’Flynn N. Diagnosis and management of patients with Lyme disease: NICE guideline. Br J Gen Pract 2018 Nov;68(676):546-547. [doi: 10.3399/bjgp18X699713] [Medline: 30361320]
17. Grignoli N, Manoni G, Gianini J, Schulz P, Gabutti L, Petrocchi S. Clinical decision fatigue: a systematic and scoping review with meta-synthesis. Fam Med Community Health 2025 Feb 8;13(1):e003033. [doi: 10.1136/fmch-2024-003033] [Medline: 39922690]
18. Hirsch AG, Herman RJ, Rebman A, et al. Obstacles to diagnosis and treatment of Lyme disease in the USA: a qualitative study. BMJ Open 2018 Jun 11;8(6):e021367. [doi: 10.1136/bmjopen-2017-021367] [Medline: 29895655]
19. Van Hout MC. The controversies, challenges and complexities of Lyme disease: a narrative review. J Pharm Pharm Sci 2018;21(1):429-436. [doi: 10.18433/jpps30254] [Medline: 30458921]

XSL•FO
RenderX

20. Johnson L, Wilcox S, Mankoff J, Stricker RB. Severity of chronic Lyme disease compared to other chronic conditions: a quality of life survey. PeerJ 2014;2:e322. [doi: 10.7717/peerj.322] [Medline: 24749006]
21. Maloney EL. Controversies in persistent (chronic) Lyme disease. J Infus Nurs 2016;39(6):369-375. [doi: 10.1097/NAN.0000000000000195] [Medline: 27755213]
22. Rebman AW, Bechtold KT, Yang T, et al. The clinical, symptom, and quality-of-life characterization of a well-defined group of patients with posttreatment Lyme disease syndrome. Front Med (Lausanne) 2017;4:224. [doi: 10.3389/fmed.2017.00224] [Medline: 29312942]
23. Munsch N, Martin A, Gruarin S, et al. Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study. J Med Internet Res 2020 Oct 6;22(10):e21299. [doi: 10.2196/21299] [Medline: 33001828]
24. Harada Y, Sakamoto T, Sugimoto S, Shimizu T. Longitudinal changes in diagnostic accuracy of a differential diagnosis list developed by an AI-based symptom checker: retrospective observational study. JMIR Form Res 2024 May 17;8:e53985. [doi: 10.2196/53985] [Medline: 38758588]
25. Levine DM, Tuwani R, Kompa B, et al. The diagnostic and triage accuracy of the GPT-3 artificial intelligence model: an observational study. Lancet Digit Health 2024 Aug;6(8):e555-e561. [doi: 10.1016/S2589-7500(24)00097-9] [Medline: 39059888]
26. Berry AC, Cash BD, Wang B, et al. Online symptom checker diagnostic and triage accuracy for HIV and hepatitis C. Epidemiol Infect 2019 Jan;147:e104. [doi: 10.1017/S0950268819000268] [Medline: 30869052]
27. Hageman M, Anderson J, Blok R, Bossen JKJ, Ring D. Internet self-diagnosis in hand surgery. Hand (N Y) 2015 Sep;10(3):565-569. [doi: 10.1007/s11552-014-9707-x] [Medline: 26330798]
28. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Aug 24;17(1):362. [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
29. Yoshida Y, Thomas Clark G. Accuracy of online symptom checkers for diagnosis of orofacial pain and oral medicine disease. J Prosthodont Res 2021 Jun 30;65(2):186-190. [doi: 10.2186/jpr.JPOR_2019_499] [Medline: 32938875]
30. Levkovich I, Elyoseph Z. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. Fam Med Community Health 2023 Sep;11(4):e002391. [doi: 10.1136/fmch-2023-002391] [Medline: 37844967]
31. Lyme Disease (Borrelia burgdorferi) 2022 Case Definition. Centers for Disease Control and Prevention (CDC). URL: https://ndc.services.cdc.gov/case-definitions/lyme-disease-2022/#:~:text=IgM%20Western%20Blot%20(WB)%20is%20considered%20positive,21%2C%2022%2C%2023%2C%2024%20or%2025%20kDA [accessed 2025-06-04]
32. Maxwell SP, McNeely CL, Brooks C, Thomas K. Triangulating the new frontier of health geo-data: assessing tick-borne disease risk as an occupational hazard among vulnerable populations. Int J Environ Res Public Health 2022 Aug 2;19(15):9449. [doi: 10.3390/ijerph19159449] [Medline: 35954807]
33. Isabel - The Symptom Checker doctors use and trust. Isabel. URL: https://symptomchecker.isabelhealthcare.com [accessed 2026-06-23]
34. MediFind. URL: https://www.medifind.com/ [accessed 2026-06-23]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
DU: diagnostic utility
EM: erythema migrans
ER: emergency room
LD: Lyme disease
NLP: natural language processing

Edited by E Mensah; submitted 10.Nov.2025; peer-reviewed by G Trevisan, H Wu, LV Rao, M Prieto; revised version received 07.Mar.2026; accepted 13.Jun.2026; published 13.Jul.2026.

Please cite as:
Maxwell SP, McNeely CL, Zeraatpisheh A, Brooks C, Maxwell C, Thomas KC
AI-Based Diagnostic Platform Capabilities With Lyme Disease as a Use Case: Integrative Exploration
Online J Public Health Inform 2026;18:e87529
URL: https://ojphi.jmir.org/2026/1/e87529
doi:10.2196/87529

XSL•FO
RenderX

© Sarah P Maxwell, Connie L McNeely, Abdollah Zeraatpisheh, Chris Brooks, Claire Maxwell, Kevin C Thomas. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 13.Jul.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Automating Diagnosis of Skin Neglected Tropical Diseases via Patient Metadata through Machine Learning Model with Adaptive Balancing and Dual Cross-Validation: Retrospective Diagnostic Accuracy Study

Yohannes Minyilu[1], MSc; Mohammed Abebe Yimer[1], PhD; Million Meshesha[2], PhD

[1]Faculty of Computing and Software Engineering, Arba Minch Institute of Technology, Arba Minch University, P.O. Box 21, Arba Minch, Ethiopia
[2]School of Information Science, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia

Corresponding Author:
Yohannes Minyilu, MSc
Faculty of Computing and Software Engineering, Arba Minch Institute of Technology, Arba Minch University, P.O. Box 21, Arba Minch, Ethiopia

Abstract

Background: Skin neglected tropical diseases (NTDs) are the most prevalent diseases worldwide, affecting people living in resource-limited areas with low health care services and trained professionals. While machine learning (ML)–based diagnostic tools can be used for initial clinical assessment and patient screening, especially in resource-limited areas (including in Ethiopia), little effort has been made in this area.

Objective: This pilot study develops a foundational ML model for the diagnosis of skin NTDs using patient metadata to analyze the feasibility of ML-based models for skin NTDs by identifying and experimentally evaluating 8 ML models.

Methods: For this study, we acquired a tabular skin NTD diagnostic dataset collected from a specific affected district in the southwest of Ethiopia. We used the data in 3 different structures, which include using the initial dataset (IDS) that contains huge null values, using a final dataset (FDS) created through preliminary preprocessing, and a third dataset created by applying feature engineering (FEFDS). Selecting 8 ML models, we trained the models in 4 major experimental settings: baseline training, handling structural missing values, handling severe class imbalance through conditional class weighting, and a hybrid approach based on robust dual cross-validation (CV) consisting of an outer repeated stratified *k*-fold and nested CV methods. We used the macro and class-specific metrics (such as precision, recall, and F_1-score), including balanced accuracy, due to the severe class imbalance. Feature importance scores are also used for evaluating overall model performance.

Results: After the final training applying the hybrid approach, 4 models scored a perfect test score (1.0) across all the metrics and all experiments except naïve Bayes and multilayer perceptron (MLP), similarly scoring 0.997 balanced accuracy, 0.97 recall, and 0.985 F_1-score. In the same experiment, the nested CV loop revealed a slight performance drop for light gradient boosting machine (LightGBM) and extreme gradient boosting (XGBoost), though both models similarly maintained higher mean scores of balanced accuracy and macro recall of 0.993 (SD 0.013), including the mean macro F_1-score of 0.996 (SD 0.007), scoring declining performance with a mean score of 0.993 (SD 0.013) in balanced accuracy, macro recall, and F_1-score, showing predictive biases. In terms of feature utilization, only CatBoost showed optimal feature ranking, while 4 models showed over-feature utilization (selection bias), with 3 models using small subsets of features (showing feature parsimony).

Conclusions: Overall, this study has been highly challenged by data scarcity, class imbalances, limited disease representation, specific geographic representation, and lack of more data modalities. Hence, further studies are suggested to confirm the results on larger datasets having a representative distribution of disease classes and geographic locations.

(Online J Public Health Inform 2026;18:e84966) doi:10.2196/84966

KEYWORDS

skin NTDs; patient metadata; ML models; feature importance; class weighting; dual cross-validation; machine learning; neglected tropical diseases

XSL•FO
RenderX

Introduction

Background

Neglected tropical diseases (NTDs) are the most prevalent diseases worldwide. According to the World Health Organization (WHO), NTDs represent about 20 different diseases, such as podoconiosis, scabies, tungiasis, buruli ulcer, cutaneous leishmaniasis (CL), leprosy, mycetoma, and rabies [1-3]. Ten of the NTDs, according to WHO, are known to exhibit primary skin symptoms (hence skin NTDs) or related clinical symptoms that often can cause bodily harm and disfigurement, which can cause disability [4]. As a tropical country, the majority of NTDs are present in Ethiopia [5], particularly in the remote areas of the country. Recent figures showed that Ethiopia remains one of the countries with a high burden of NTDs, where more than 75 million people are at risk of contracting at least one of the NTDs [6,7].

Clinically, different diagnostic procedures are used for the diagnosis of skin NTDs, where the dermatological approach is the primary method used. Marks et al [8], on behalf of the WHO Diagnostic Technical Advisory Group (DTAG), highlighted that the integration of clinical and laboratory diagnostic services is the fundamental approach for skin NTDs, where initial clinical assessment of patients, followed by confirmatory tests (laboratory diagnostics), are the 2 critical steps toward enhanced diagnostics of skin NTDs. However, establishment of laboratory facilities in most NTD-vulnerable areas (based on current real-world scenarios) is hardly feasible due to higher resource requirements. Such scenarios guide efforts to look for alternative, technology-assisted solutions through decentralized diagnostic frameworks. Tadesse et al [9], who studied the decentralized diagnosis of CL in Ethiopia, claimed that with minor adjustments to health care facilities and continuous training of primary health care professionals, the diagnosis and treatment of skin NTDs can be decentralized to the health care center levels.

Furthermore, recent studies such as [6,10] suggested the feasibility of mobile-based diagnostic platforms for resource-limited settings. The authors in [11] and [12] used advanced AI-based diagnostic tools based on machine learning (ML) and deep learning (DL) methods for the diagnosis of skin NTDs. The use of such tools would facilitate the creation of localized diagnostic platforms, as they provide the opportunity for patients to get health care services around their living areas, especially in resource-limited areas. This triggers the need to investigate and implement NTD diagnostics based on effective point-of-care tools for skin NTD diagnosis [8,9,13]. Accordingly, ML-based diagnostic tools based on clinical patient data can be used for initial clinical assessment of patients and patient screening, as well as patient record organization and management, especially in resource-limited areas. Most importantly, in resource-constrained areas, ML-based diagnostic tools implemented based on the differential diagnostic approach can be used by middle-level health care workers for early diagnosis of patients with skin NTDs. Therefore, this study investigates the feasibility of an ML-based skin NTD diagnostic framework by proposing an ML-based diagnostic model for skin NTDs using patient metadata. Accordingly, the study is designed to be a pilot study that intends to establish the foundational basis for the realization of an intelligent diagnostic model for skin NTDs.

In this pilot study, we establish a robust and integrated framework for the systematic investigation of the feasibility of high-performing diagnostic ML models for skin NTDs based on tabular clinical data. This includes establishing integrated pipelines for data preprocessing, handling missing data, and addressing class imbalance, including the identification, synthesis, and evaluation of ML architectures. For this study, we used a novel dataset created using newly collected tabular patient metadata of 3 skin NTDs (podoconiosis, scabies, and tungiasis) to train and evaluate 8 selected ML models. We further conducted a comparative performance analysis of the trained models to identify the model with the highest performance and reliability, as well as the optimal ML strategy that resulted in the best performance. The study also analyzed the impact of missing values and strategies for handling missing data (imputation methods), which serves as a benchmark for further studies. However, as the study presents a foundational framework, it does not intend to build a full-fledged and deployment-ready diagnostic model; as such, complete systems require sufficiently large datasets, high computational resources, complex and multistaged evaluation, and regulatory procedures before deployment.

Overall, apart from being the first study that proposes the application of ML-based diagnostics for skin NTDs using tabular data, this study presents other novel contributions, which include presenting a novel tabular skin NTD dataset and the establishment of a harmonized ML pipeline. Accordingly, we identified and used a set of ML strategies and created a robust ML model development pipeline for the systematic application, evaluation, and analysis of methods and models, which include comprehensive baseline training setup, rigorous feature ranking and importance analysis, robust implementation of missing data handling, and the harmonized approach—the robust ML pipeline that synergistically integrates conditional class weighting with a dual CV evaluation approach based on the integration of 2 CV methods (repeated and nested CV methods). The collective implementation of all these strategies helps in establishing an overall robust ML pipeline, clearly marking the novelty of our study.

Related Works

Machine Learning Models for Structured Data

Developing ML-based diagnostic tools requires carefully devised strategies for the selection, utilization, evaluation, and optimization of the ML models and tools. Accordingly, the type of data to be used, computational resources available for model building and deployment, expected model performances, undesired model behaviors such as overfitting, model optimization strategies to be used, and the nature of the disease to be diagnosed are the major parameters that determine the ML methods and strategies to be used. For small-sized datasets, support vector machines (SVMs), decision trees, and random forests (RF) are the preferred ML algorithms, while the convolutional neural network (CNN)–based DL models can

XSL•FO
RenderX

also be used [14,15]. However, the tree-based model extreme gradient boosting (XGBoost) is the most suitable model for tabular data since it requires fewer hyperparameter tuning tasks compared to the DL-based models, as stated in [16]. Generally, the gradient-boosted tree models of XGBoost, light gradient boosting machine (LightGBM), and CatBoost are well-suited for tabular data, although they can be challenged by model complexity [17], higher computational resource requirements such as GPUs for large datasets [18], and possible issues during real-time deployment on edge devices [19].

Machine Learning Methods for Data-Related Constraints

There are multiple ML strategies and models for handling dataset-related issues, as it is sometimes challenging to acquire complete datasets that are sufficient to train ML and DL models. For small-sized tabular datasets, traditional ML models and ensemble approaches are the most suitable methods due to their stability and interpretability, as indicated in recent studies. Accordingly, the tree-based ensemble models of RF and extra trees are most often the top-performing models for the classification of diseases in dermatology based on tabular patient metadata, including the situation of having small-sized tabular datasets [20]. In terms of performance reliability, *k*-nearest neighbor (KNN) and SVM are the most stable models that resulted in superior performance compared to complex neural network–based models for small-sized structured clinical datasets [21,22]. While the boosted tree-based model XGBoost requires a larger number of samples to reach performance stability, the RF model appears to be a more reliable model for small-sized datasets [22]. Regarding missing data, studies suggested that the gradient-boosted tree models of CatBoost, XGBoost, and LightGBM are highly robust to missing data when trained on incomplete patient records. Accordingly, CatBoost (with nan_mode enabled) is the most robust model, as it has an internal mechanism to treat “NaN” values as a separate split direction [23]. The skin disease detection study by Aquil et al [20] highlighted that XGBoost and LightGBM models have the ability of learning “default direction” for missing values during training. Accordingly, if a given symptom (feature) has a null value, the model automatically assigns this feature to the branch that minimizes the loss based on other patients.

Machine Learning Methods for Skin NTD Diagnosis

Most previous studies conducted for the diagnosis of skin NTDs heavily rely on skin images. Accordingly, the previous studies [23-25] used DL-based methods to apply diagnostic models for skin NTDs, where all these studies used only skin image-based approaches. On the other hand, previous works that implemented ML methods for the diagnosis of skin NTDs using only structured clinical data were difficult to find, except 2 studies that used tabular patient metadata along with skin image data. Accordingly, the study by Barbieri et al [11] used structured clinical data of leprosy patients to integrate with skin lesion images and used ML algorithms that include RF and XGBoost for handling the clinical patient data in their multimodal data fusion study. Similarly, Achary et al [12] used tabular clinical metadata to apply multimodal data fusion by combining the metadata with skin images of patients using generative models for text data augmentation. In both studies, the tabular clinical data of patients are not used independently for skin NTD diagnosis; instead, they are merged with image data based on multimodal data fusion methods.

Overall, the results of the literature review underscored that structured clinical metadata of patients are not fully used, let alone the independent utilization, demonstrating the development of ML-based diagnostic models for skin NTDs based on tabular clinical patient data, indicating a clear research gap. One of the biggest challenges in building AI-based diagnostic tools for skin NTDs is the lack of large-scale, complete, and properly annotated datasets, as affirmed by the study by Achary et al [12]—the study used generative models to generate textual data to fill the gap of data scarcity. Conversely, the selection and use of specific ML strategies require further investigation, given that skin NTD data are naturally scarce resources, including limited previous studies for comparative analysis of methods. All these factors clearly mark research-wide gaps that forced the authors of this study to conduct this study.

Methods

Study Design

This pilot study develops a foundational benchmark diagnostic model for skin NTDs using structured (tabular) clinical patient data by selecting the best-performing model from a set of 8 ML models initially selected for the experiments. For a complete and fair comparison, all models are equally trained on similar datasets, using similar training settings. To achieve this objective, the study relies on capturing and analyzing model performance results by conducting a series of model training experiments. This highlights both the qualitative and quantitative aspects based on the experimental approach. Therefore, this study uses a mixed research strategy. However, this study does not currently intend to realize a fully deployable diagnostic model.

Ethical Considerations

This study uses a novel dataset created by acquiring clinical records consisting of demographic information and disease-specific diagnostic data of patients with skin NTDs. The entire data collection and acquisition processes are carried out in a professional and ethical manner, which include acquiring an ethical clearance letter that authorizes the data collection and use of the collected data for this study, and confirming all legal and ethical issues, including written consent of patients. To achieve this, the authors of this study have acquired a proper ethical clearance letter from the Institutional Review Board committee at Arba Minch University (approval protocol number: YM23161).

Data Acquisition and Dataset Preparation

Data Source

For this study, we created a new hand-crafted dataset using the data collected from patients with skin NTDs living in Gacho Baba district of the Gamo zone, southwest of Ethiopia, which is one of the remote areas representing underserved communities

XSL•FO
RenderX

in terms of health care services. The data were collected by a team of researchers from the Collaborative Research and Training Center for NTDs, College of Medicine and Health Sciences of Arba Minch University, in a project-based research conducted for the assessment of skin NTD burden. For this study, the original (first-hand) data were acquired through institutional collaboration (acknowledgments section) and used to achieve the objective of this study.

Initial Diagnostics and Confirmation

The entire data collection was initially conducted in a large-scale community-based mass drug administration (MDA) campaign in the specified community based on on-site clinical diagnostic procedures. At the MDA site, however, extended medical resources like laboratory facilities, specialized medical imaging, or any advanced medical infrastructure were not available. In such settings, the initial clinical diagnostic procedures were performed by trained frontline health care workers (mostly middle-level clinicians and nurses) who were actively involved in the day-to-day clinical diagnostics of patients around the area, and a professional dermatologist from Arba Minch General Hospital who participated during the MDA campaign. Therefore, the on-site diagnostics were performed entirely based on visual clinical examination procedures by the frontline clinicians, with strict diagnostic confirmation by the professional dermatologist.

As the data collection involved actual field-based diagnoses, the establishment of initial ground truth was challenged by limitations related to field logistics. To overcome this challenge, a hierarchical diagnostic verification protocol was applied by the data collection team to clinically confirm the diagnostics. Hence, each of the initial diagnostic procedures underwent a mandatory clinical review by a professional dermatologist from Arba Minch General Hospital to confirm the initial diagnostics. While this procedure confirms the correct classification of the skin NTD diagnosed based on the signs and symptoms presented by the patients, it also helped in creating higher-level diagnostic fidelity. Overall, the ground truth we used for the classification of the skin NTDs for this study complies with the WHO-standardized clinical case definitions [26]. After all these clinical and ethical procedures were completed, the authors of this study used the clinically validated data for the ML-based research in this study after acquiring the proper authorization through ethical clearance.

Data Collection

Overall, the data collection was conducted based on a community-based screening approach for assessing the burden of skin NTDs, which included on-site examination, treatment, and registration of patient information. To achieve this, the data collection team used a well-prepared data collection and patient registration form, which was prepared by consulting professional dermatologists from Arba Minch General Hospital. The data collection involved 4029 participants during the initial screening and diagnosis, targeting the identification and diagnosis of 8 skin NTDs that are endemic to Ethiopia, which include mycetoma, leprosy, CL, lymphatic filariasis, podoconiosis (podo), onchocerciasis, tungiasis, and head lice. Finally, according to the data collection report, the specified community was highly affected by 3 of the skin NTDs, namely scabies, tungiasis, and podoconiosis. Thus, we selected these 3 diseases for our study based on the prevalence and availability of sufficient data. The other 5 mentioned diseases have very little (or no data at all) in the specified community at the time of data collection; hence, they are not used in the final dataset.

Limitations of the Data Collection

Generally, collecting large-scale skin NTD data from wider geographic areas (like Ethiopia) with complete representation of all endemic diseases is a complex and highly resource-intensive process, which is almost impossible to achieve in a single research project. Hence, our study is limited to a single geographic location in the Gacho Baba district, southwest of Ethiopia. This confinement to the specific location limited the number of endemic diseases to only 3 skin NTDs (scabies, tungiasis, and podoconiosis) out of the 8 overall endemic diseases, which forced us to exclude other skin NTDs like buruli ulcer, leishmaniasis, leprosy, and mycetoma. However, as a preliminary pilot study, we used the acquired skin NTD data to build a benchmark skin NTD diagnostic model, with the potential of extending the disease representation through further efforts.

Dataset Creation

During initial collection, the patient data collected for each of the 3 diseases were recorded separately using individual Microsoft Excel files for each disease. Hence, we initially created 3 separate datasets, which represent (1) scabies, which consists of 955 instances having 10 features (6 demographic and 4 disease-specific); (2) tungiasis, consisting of 474 instances having 8 features (6 demographic and 2 disease-specific); and (3) podoconiosis, comprising 66 instances having 11 features (6 demographic and 5 disease-specific), all before data preprocessing. In each of these 3 separate datasets, all disease-specific features (symptoms) are represented in binary categorical form using either “1” or “0” to signify the presence or absence of a given disease-specific symptom in a patient, which were applied after clinical confirmation of all diagnoses. For instance, “slowenlarging_swelling(feet&legs)” is a symptom specifically assessed while diagnosing podoconiosis, and all 66 podo patients were assessed if they manifest this particular symptom. Hence, after final clinical confirmation of the diagnosis, this feature is encoded as either “1” if a given patient showed this symptom or “0” otherwise, where a similar pattern was used for all symptoms of the 3 diseases.

Finally, we handcrafted our initial dataset (named it IDS) by merging the 3 individual disease-specific datasets. Originally, IDS comprises a total of 1495 instances with 17 trainable features (6 demographic and 11 diagnostic) and the target variable “disease_diagnosed” defining the class labels. After initial data preprocessing, IDS is transformed into a new structure having 1495 instances with 28 trainable features (17 demographic and 11 diagnostic) plus the target variable. However, the merger of the 3 separate disease-specific datasets introduced a different challenge due to the disease-specific symptoms, as not all symptoms are applicable to all diseases—a specific symptom that can be used to diagnose scabies is not applicable to the other 2 diseases. This created 3 blocks of missing values in IDS, showing a structural missingness issue

XSL•FO
RenderX

[27]. Hence, using ML-based operations, we created 2 additional datasets through transformation: first, we created a final dataset (FDS), having 28 trainable features (17 demographic and 11 disease-specific features); next, we created a different dataset by applying feature engineering on the final dataset (FEFDS), having 38 trainable features (17 demographic and 11 disease-specific features, including 10 symptom applicability indicator flags introduced after feature engineering). We created all these 3 different datasets in 3 separate CSV files to address structural missingness, ensure reproducibility, and maintain dataset integrity across different experimental settings, thereby allowing us to establish a consistent training and evaluation pipeline.

Exploratory Data Analysis

Dataset Characteristics and Missingness Analysis

Multimedia Appendix 1 presents the overall distribution of instances (n=1495), disease classes, and missing values in IDS. As shown in Figure S1 of Multimedia Appendix 1, the scabies disease class represents the largest proportion with a total of 955 (63.9%) instances, while tungiasis represents the second largest proportion with 474 (31.7%) instances, and podoconiosis has only 66 (4.4%) instances. Gender-wise, 66.6% (995/1495) of the patients are male, while 33.4% (500/1495) are female. These statistics clearly highlight that the dataset exhibits a severe class imbalance among the 3 disease classes, having an extremely distorted distribution with an overall imbalance ratio of 14.5 between scabies (majority class) and podoconiosis (minority class).

Correspondingly, Figure S2 in Multimedia Appendix 1 depicts the statistical summary of missing values in the dataset. As stated, our new dataset is created by merging 3 separate disease-specific datasets, each of which has a unique set of clinically relevant symptoms that are not applicable to the other diseases. For instance, the symptom “PPSEs” (which represents “papule, pustule, scratches, and excoriations”) is assessed only for scabies patients, where this field is structurally empty for podoconiosis and tungiasis patients. Ultimately, this created 3 blocks of missing values that define the case of structural missingness or missing by design [27], where our new dataset contains a total of 11,347 such structural null values. The missing values in our dataset are not random omissions created as a result of the random absence or exclusion of symptoms that are called “missing completely at random (MCAR)” or “missing at random (MAR)” [28]. Overall, as depicted in Multimedia Appendix 1, the scabies class has the highest number of null values, with 58.9% (6685/11,347) of the total number of null values (having 7 null values per each of the 955 rows), while tungiasis has 37.6% (4266/11,347), and podoconiosis has 3.5% (396/11,347) since it has the least number of instances.

Feature Description

Table 1 presents the overall summary of our first dataset, IDS, including its overall statistical summary, including its overall statistical summary of class distribution, feature categories, and the initial feature sets of IDS.

As shown, our new dataset (IDS) primarily contains two major categories of patient data: (1) demographics, which include 2 numerical features (age and weight) and 3 categorical features (gender, marital status, educational level, and occupation) of the patient information, and (2) clinical data, which comprises 11 disease-specific symptoms from the 3 classes. Class-wise, podoconiosis has 5 specific features (disease-specific symptoms) that are not applicable to the other 2 diseases, and the scabies class has 4 unique symptoms applicable only to scabies, while there are only 2 tungiasis-specific symptoms (features). Overall, our dataset has 2 features representing numerical data (age and weight), while the remaining features are categorical, which in turn include 3 ordinal categorical features, representing sex, educational level, and one disease-specific feature (the jigger infestation classification feature of tungiasis); 2 nominal categorical features (marital status and occupation), representing the unordered categorical data that include marital status occupation; and 10 binary categorical features, represented as “1” or “0” indicating the presence or absence of a given disease-specific symptom. Table S1 in Multimedia Appendix 2 presents the complete dataset information.

XSL•FO
RenderX

Table . Summarized statistical description of the new dataset used for this study.

Features	Podoconiosis (n=66)	Scabies (n=955)	Tungiasis (n=474)	*P* value
Demographic				
Age (years), mean (SD)	44.68 (13.59)	17.81 (15.17)	21.90 (20.28)	<.001
Sex, n (%)				.104
Female	30 (45.5)	316 (33.1)	154 (32.5)	
Male	36 (54.5)	639 (66.9)	320 (67.5)	
Weight (kg), mean (SD)	52.02 (7.73)	33.31 (15.74)	32.47 (16.70)	<.001
Educational level, mean (SD)	1.26 (0.54)	2.34 (1.38)	2.37 (1.61)	<.001
Clinical				
Disease-specific clinical symptoms	5 (specific symptoms) • Slow-enlarging swelling • History of barefoot • Living podo in endemic area • Acute attacks • Involving bilateral legs	4 (specific symptoms) • Night-worsened itchy lesion • Family history of similar condition • PPSEs[a] • FWABBP[b]	2 (specific symptoms) • Body part infested (hand/leg) • Jigger infestation classification	

[a]Papule, pustule, scratches, and excoriations.

[b]Finger webs, axilla, breast, buttock, and penis.

Data Preparation and Preprocessing

Data Labeling and Initial Data Preprocessing

In this study, we used IDS in 3 different settings. First, IDS is used with minimal preprocessing operations for developing baseline models. This dataset is used to mimic the original real-world scenario, which is sometimes characterized by incomplete datasets having multiple missing values. As a base dataset, the majority of the essential data preprocessing operations are applied to IDS, while we applied only normalization for FDS and FEFDS, as these datasets were derived from the feature-encoded IDS. Overall, for data preprocessing, we applied data encoding, which mainly included data encoding using label encoding and one-hot encoding methods, including normalization.

For encoding the ordinal categorical features ("sex," "educational_level," "JI_classification(#Jig-Inf)," including the target variable "disease_diagnosed"), the Label Encoder method is used. As the remaining 12 features represent nominal categorical data, which include "marital_status," "occupation," and 10 disease-specific symptoms, we applied the one-hot encoding method to transform the data under these features. Data normalization was the other preprocessing operation we performed. Accordingly, the numerical features of age (with minimum 2 and maximum 77) and weight (with minimum 8 and maximum 80) are normalized using the "standard scaler" technique to bring these features on the same scale with the other features. This allowed us to maintain a standard normal distribution with zero mean and unit variance within the dataset [29].

Handling Structural Missing Data: Feature Engineering

As stated, our new dataset (IDS) contains a total of 11,347 structural null values, which are not random omissions created as a result of the random absence or exclusion of symptoms that are called MCAR or MAR [28]. Over reliance on conventional imputation techniques such as simple imputation or using other methods like mean, median, or multivariate imputation by chained equation (MICE) would introduce biases that could potentially cause both statistical and clinical issues [27,28], having contradictory implications: positively, imputing missing values with "0" values is logically correct, easy to apply, and does not cause any loss of information; while negatively, imputing the missing values with "0" values will introduce hidden danger if a patient develops a secondary (additional) symptom due to a different cause. In our case, for instance, if our models are trained using the fact that "if the values of the 'Slowenlarging_swelling(feet&legs)' column are non-zero then, the disease cannot be scabies or tungiasis." However, such restrictions will affect the generalizability of our models when dealing with patients having mixed symptoms if a patient also has swelling on the feet due to different causes. Therefore, this requires careful analysis and selection of strategies to be used for the imputation, which adds robustness to our models.

Therefore, we created the final dataset by applying a hybrid data encoding strategy: (1) we encoded all symptoms as binary, putting 1 if a given symptom is present (applicable) to a particular disease and putting 0 otherwise; and (2) we introduced a "symptom missingness indicator" parameters, which flags 1 if a given symptom applies to a particular disease and 0 if that symptom does not apply. Therefore, we created a new column called "DISEASE-SYMPTOM_applicable" as a missingness indicator, indicating if that "SYMPTOM" is applicable to the "DISEASE" or not. For instance, since "PPSEs" is a symptom

XSL•FO
RenderX

applicable only to scabies and not applicable to podoconiosis or tungiasis, the "PPSEs_applicable" column for all scabies patients is encoded as "1" whether a patient has "PPSEs" or not, while this column is encoded as "0" for podoconiosis and tungiasis patients. The repeated application of this method across 10 disease-specific features adds 10 additional confirmatory columns, which expanded the number of trainable columns of the dataset to 38.

Experimental Setup

Overall Experimental Configurations

This study is conducted based on the stratified *k*-fold cross-validation (CV) method to evaluate possible performance variations resulting from changes in the dataset and ML models and methods. The hold-out method is excluded from the experimental setup due to high-level overfitting recorded during preliminary experiments. Therefore, we conducted extensive experiments using our new dataset in 3 structures (IDS, FDS, and FEFDS) with an 80/20 train-test split based on the stratified 5-fold CV method. We designed 4 separate experimental settings that include (1) baseline training, where IDS (the original dataset) is used with minimal data preprocessing (only data encoding); (2) handling structural missingness, which includes 2 separate training experiments conducted for handling the structural missing values—the first experiment includes training using FDS that applies the simple imputer method, while the next experiment was conducted on FEFDS that applies the symptom applicability indicator flags created after feature engineering—each of these two experiments exactly reflects the other except for the datasets used, aiming to comparatively evaluate performance impacts between the two methods; (3) conditional class weighting; and (4) the hybrid approach, the experiment that harmoniously combines class weighting and the robust dual validation methods, all applied on the dataset created using feature engineering (experiment 3).

Conditional Class Weighting

To address the severe class imbalance, we applied a class weighting method. Due to the difference in the internal nature of one of our selected models in accepting the "sample_weight" parameter of the scikit-learn library [30], we conditionally applied the class weighting method. Therefore, for all 7 models that inherently accept the "sample_weight" parameter, these sample weights are automatically computed using the built-in method with the "class_weight" option set to "balanced," then directly passed to their fit methods. Conversely, as the multilayer perceptron (MLP)–based model does not accept the "sample_weight" parameter, we implemented a fallback strategy based on resampling to replicate the instances of the minority class. In this setting, instances from the training set are resampled with replacement based on probabilities equal to the sample weights computed at the beginning of the class weighting method. We applied this method to facilitate the consistent application of the class weighting method across all the selected models for a complete and fair performance comparison.

The Robust Dual Validation Framework

Given the size and distribution of our dataset, using a single *k*-fold (5-fold) split might not help to critically assess the stability and reliability of the models, where the repeated stratified *k*-fold CV method plays vital roles in overcoming performance instability issues [29,31]. Therefore, to establish robust and unbiased performance estimates, we harmoniously applied the repeated and nested stratified CV methods.

Repeated Stratified *k*-Fold CV (Outer Loop)

For a severely imbalanced dataset like ours, the distribution of classes across each fold is preserved through stratification. Accordingly, the repeated stratified CV method allows each fold to hold a consistent distribution among the 3 disease classes as a result of the repeated shuffling [32]. Hence, to critically assess the average performance and stability of selected models, we applied the repeated stratified *k*-fold CV method by repeating our previous 5-fold CV loops 10 times to randomly repartition the training set into 5-folds across each repeat while maintaining the class distribution.

Nested CV (Inner Loop)

The nested stratified CV is the other ML-based strategy we applied in this study, which is a dual validation strategy consisting of 2 main loops, an inner loop for hyperparameter tuning and an outer loop for validation, collaboratively operating to achieve unbiased performance estimates [30,33]. Accordingly, we implemented the nested CV that uses a stratified *k*-fold loop for the outer loop with 5 folds and another loop for the inner loop with 3 folds. The inner folds are sequentially used by each of the outer folds for selecting the best hyperparameters by running the grid search (GridSearchCV) method on the training portion of each of the outer folds. To ensure unbiased performance estimates, we implemented a procedure to evaluate the best model on the outer test fold, where all the scores on the outer test fold are aggregated.

Overall, in this strategy, we applied the stratification method to maintain the proportional representation of the instances of the minority class (podo) across all training and testing phases. Additionally, to reduce data partitioning bias and evaluate the stability of model prediction results, the complete nested process is repeated 10 times (the repeated stratified *k*-fold CV with *k*=10 is applied) using different random seeds. In this setting, the conditional class balancing method is applied inside the inner loop (inside the inner training folds).

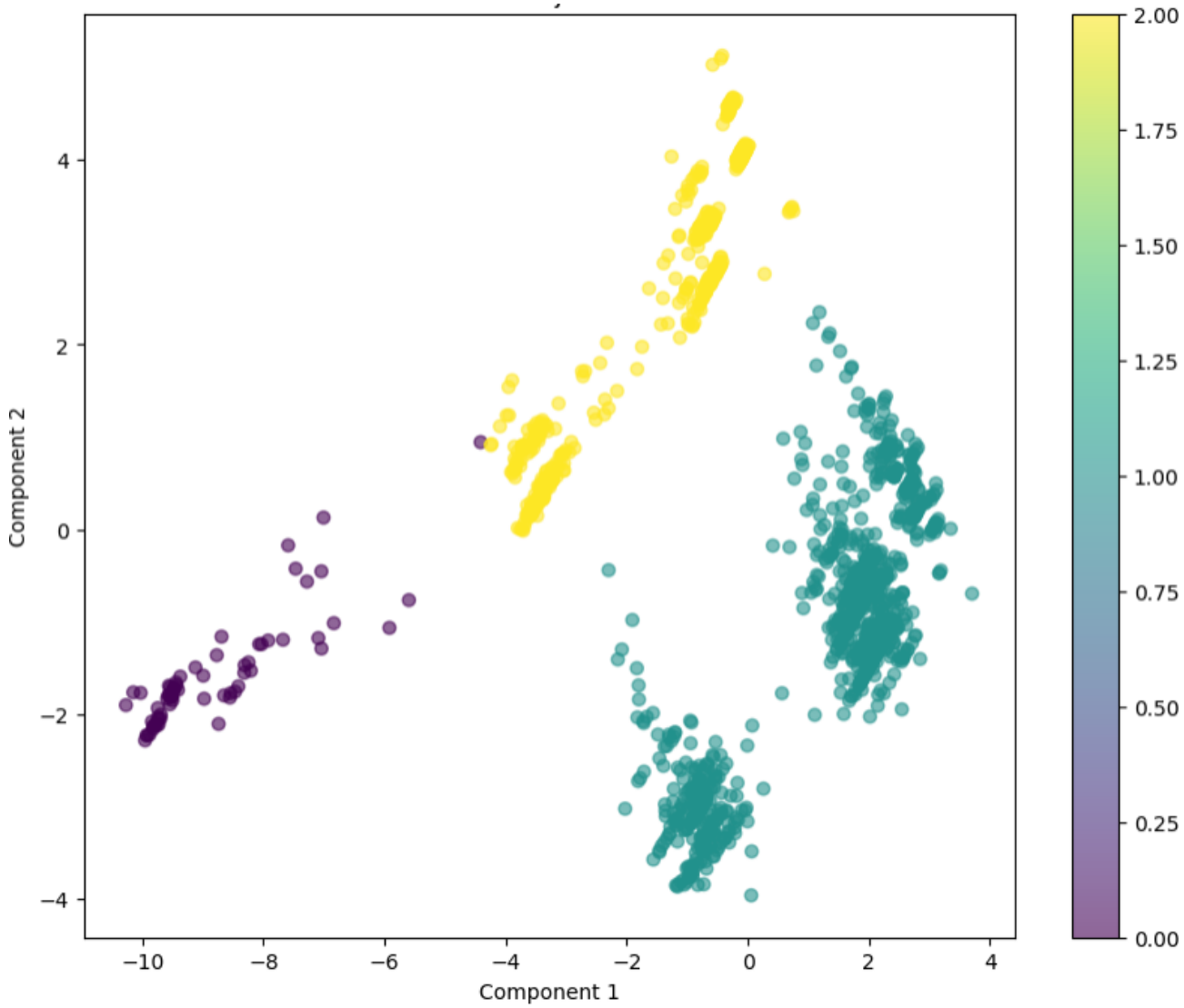
Model Development

Model Selection

This study uses datasets having 1495 instances with only 38 trainable features, where all the features are mostly linear, as shown by the principal component analysis (PCA) projection in Figure 1.

XSL•FO
RenderX

Figure 1. The PCA projection analysis of the features of the datasets. PCA: principal component analysis.



As shown, the features of the 3 disease classes are highly linear, which shows the tendency of the features to be linearly separable. Hence, given the PCA projection analysis and the datasets to be used, 6 parameters are used to select the models. The parameters used include the model's ability in handling small-sized datasets, the nature of the data used (mostly linear), the robustness of the model in handling higher-level class imbalances, the robustness of the model in handling data with missing values, higher model performance expected, and model simplicity and interpretability for future deployment of the models on portable devices. Therefore, based on the extended analysis of size, complexity, and overall nature of our datasets, ML models, and the intent of the study, we selected 8 ML models: 3 nontree models, namely naïve Bayes (NB), SVM (with linear kernel), and a simple CNN-based model using MLP; and 5 tree-based models, which include RF (ensemble tree-based model) and 4 boosted tree models of CatBoost, AdaBoost, LightGBM, and XGBoost.

Model Evaluation Methods

For this study, we applied a multifaceted evaluation strategy, given the nature of our new dataset. First, the study entirely relies on the stratified *k*-fold CV method, as it allowed us to achieve more robust and lower-variance estimates. Hence, we applied a 5-fold (k=5) stratified *k*-fold CV method, as the hold-out method resulted in total overfitting during preliminary experimental trainings. For this study, using the standard accuracy might be misleading given the small-sized dataset used, the simplicity or complexity of models with respect to the dataset used, and the tendency of models to favor the majority class due to the extreme class imbalance [34]. To address this, we used "balanced accuracy," which is a metric used to reliably evaluate models by giving equal weights for all classes by averaging sensitivity (recall) and specificity (ie, rate of correctly predicted negative classes) of the models [28,29].

The study highly focuses on the F_1-score metric to balance between the precision and recall. The G-mean score, which represents the geometric mean of TPR (true positive rate or recall) and TNR (true negative rate or specificity) [35], is used to evaluate the class-wise accuracy of the models on the 3 disease classes. We used the G-mean along with AUPRC (area under the precision-recall curve) as these metrics are not affected by class imbalances [36], where the AUPRC metric is used to robustly assess the models' predictive performance on the minority class. To evaluate the models' class-wise separation

XSL•FO
RenderX

performance, the ROC AUC metrics are computed based on the one-vs-one and one-vs-rest classification approaches. Apart from the macro average scores, the class-specific performance of each model is also evaluated to assess how well each model performs on each class, given the severe class imbalance. Hence, the class-specific precision (positive predictive value), recall (sensitivity), specificity (true negative rate), F_1-score, and NPV (negative predictive value) are used to individually assess the models on each class. Confusion matrices are heavily used to compute and visualize the class-specific scores for each model across all experimental settings. Further, we used the feature importance (FI) scores of each model to evaluate model performance and stability, monitor overfitting due to imbalance, include clinical interpretability, and identify noise and redundant features in the dataset. To apply the FI evaluation, we implemented 3 different but complementary techniques that include using built-in importance for the tree-based models, absolute coefficient magnitude for the linear model (SVM), and permutation importance for MLP and NB. Finally, based on the detailed evaluation of the models' performance, we conducted deeper analysis using descriptive statistical metrics, which include computing average or mean (μ) values, computing and evaluating numerical variations or differences (Δ) of models' performance, and analyzing performance stability of the models using SD or σ, across the different experimental settings.

Results

For this study, we conducted the training experiments based on 4 logical experimental settings, as presented by the results in the following subsection.

Baseline Training

During the first training using IDS based on the 5-fold CV method, all the selected nontree models (NB, SVM, and MLP) and AdaBoost (tree-based model) failed to train due to the null values in the dataset and terminated the training with error messages. These results verified that these 3 nontree and AdaBoost models were highly sensitive to missing values [37], due to the models' inherent mathematical paradigms of reliance on the availability of complete data compared to the tree-based models [38]. Accordingly, the 4 tree-based models—RF, CatBoost, LightGBM, and XGBoost—were able to classify the diseases with evaluation results, as presented in Table 2.

Table . Overall model training results on the initial dataset using the *k*-fold cross-validation method.

Model	Accuracy		Macro average						
	Overall accuracy	Balanced accuracy	Precision	Recall	F_1-score	G-mean	ROC-AUC[a] (OVO[b])	ROC-AUC (OVR[c])	AUPRC[d]
Random forest	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
CatBoost[e]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
XGBoost[f]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
LightGBM[g]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

[a]ROC-AUC: area under the receiver operating characteristic curve.
[b]OVO: one-vs-one.
[c]OVR: one-vs-rest.
[d]AUPRC: area under the precision-recall curve.
[e]CatBoost: categorical boosting.
[f]XGBoost: extreme gradient boosting.
[g]LightGBM: light gradient boosting machine.

As shown in Table 2, all the 4 models achieved perfect scores (1.0) in accuracy and across all the macro average metrics of precision, recall, F_1-score, G-mean, and ROC-AUC. Class-wise, all these 4 models similarly scored perfect scores (1.0) across all metrics, as shown in Table 3.

Table . Class-specific results of the 4 trainable models on IDS[a] using the *k*-fold cross-validation method.

Class	Precision	Recall	F_1-score	TNR[b]	NPV[c]
Podo	1.0	1.0	1.0	1.0	1.0
Scabies	1.0	1.0	1.0	1.0	1.0
Tungiasis	1.0	1.0	1.0	1.0	1.0

[a]IDS: initial dataset.
[b]TNR: true negative rate or specificity.
[c]NPV: negative predictive value.

XSL•FO
RenderX

Handling Structural Missingness

Table 4 presents overall model performance across the 2 experiments, demonstrating methods for handling missing values.

Starting from the second phase, all 8 models (including NB, SVM, MLP, and AdaBoost that were previously unable to train) were able to run properly as a result of replacing the missing values using the simple imputation method. However, as the null values are structural, an additional experiment demonstrating feature engineering techniques by creating missingness indicator flags was conducted to identify the suitable method for our dataset. Hence, during this experimental setting, we conducted two separate experiments demonstrating the methods we applied to handle structural missingness: (1) experiment A, which was conducted by applying the simple imputation method for the missing disease-specific values, and (2) experiment B, which applied a feature engineering technique and introduced missingness indicator flags for the missing disease-specific values. Table 4 presents the overall results recorded in these 2 trainings. As shown in Table 4, with the use of the simple imputation method, 7 models showed perfect scores (1.0) across all evaluation metrics of balanced accuracy and macro average scores of precision, recall, F_1-score, G-mean, and AUPRC scores. Only MLP scored lower performance compared to the other models by achieving 0.90 in balanced accuracy and macro recall, and 0.77 in precision and F_1-score, while achieving 0.93 G-mean and approximately 1.0 (0.999) AUPRC scores.

Table . Model performance results when trained using the *k*-fold cross-validation method.

Model	Experiment A: simple imputing						Experiment B: use of missingness indicators					
	Balanced accuracy	Precision	Recall	F_1-score	G-mean	AUPRC[a]	Balanced accuracy	Precision	Recall	F_1-score	G-mean	AUPRC
Naïve Bayes	1.0	1.0	1.0	1.0	1.0	1.0	0.974	0.996	0.974	0.985	0.986	0.972
SVM[b] (linear)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
MLP[c]	0.90	0.77	0.90	0.77	0.93	0.999	0.997	0.956	0.997	0.974	0.997	0.999
Random forest	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
AdaBoost[d]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
CatBoost[e]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
XGBoost[f]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
LightGBM[g]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

[a]AUPRC: area under the precision-recall curve.

[b]SVM: support vector machine.

[c]MLP: multilayer perceptron.

[d]AdaBoost: adaptive boosting.

[e]CatBoost: categorical boosting.

[f]XGBoost: extreme gradient boosting.

[g]LightGBM: light gradient boosting machine.

Regarding class-specific performance, MLP scored the worst per-class precision of 0.31 for podo, while scoring 1.0 for scabies and tungiasis, as shown in Table 5. MLP also has the lowest F_1-score for podo (0.47), while scoring higher for tungiasis (0.83) and an almost perfect score (0.997) for scabies. However, MLP scored the highest recall and NPV scores for podo compared to scabies and tungiasis.

XSL•FO
RenderX

Table . Model performance results when trained using the *k*-fold cross-validation method.

Class	NB[a], SVM[b], RF[c], AdaBoost[d], CatBoost[e], XGBoost[f], LightGBM[g]					Per-class scores: MLP[h]				
	Precision	Recall	F_1-score	Specificity	NPV[i]	Precision	Recall	F_1-score	Specificity	NPV
Podo	1.0	1.0	1.0	1.0	1.0	0.31	1.0	0.470	0.899	1.0
Scabies	1.0	1.0	1.0	1.0	1.0	1.0	0.995	0.997	1.0	0.991
Tungiasis	1.0	1.0	1.0	1.0	1.0	1.0	0.705	0.827	1.0	0.879

[a]NB: Naïve Bayes.
[b]SVM: support vector machine.
[c]RF: random forest.
[d]AdaBoost: adaptive boosting.
[e]CatBoost: categorical boosting.
[f]XGBoost: extreme gradient boosting.
[g]LightGBM: light gradient boosting machine.
[h]MLP: multilayer perceptron.
[i]NPV: negative predictive value.

In experiment B, all models scored similar results as the previous simple imputation experiment, except NB and MLP, as shown in Table 4. With the use of the missingness indicators, the NB model achieved declining performance by achieving 0.974 in balanced accuracy, including the 0.97 and 0.985 macro recall and F_1-scores. Conversely, MLP scored significantly improved results compared to the previous experiment with simple imputation, while the remaining 6 models scored perfect scores across all metrics. Class-wise, as shown in Table 6, while MLP achieved improved per-class scores compared to the previous per-class scores (Table 5), NB scored declining performance in the class-specific scores of recall (podo: 0.92), F_1-score (podo: 0.96; tungiasis: 0.995), specificity (tungiasis: 0.995), and NPV (podo: 0.997).

Table . Model performance results when trained using the *k*-fold cross-validation method.

Class	Naïve Bayes					MLP[a]				
	Precision	Recall	F_1-score	Specificity	NPV[b]	Precision	Recall	F_1-score	Specificity	NPV
Podo	1.0	0.92	0.96	1.0	0.997	0.87	1.0	0.93	0.993	1.0
Scabies	1.0	1.0	1.0	1.0	1.0	1.0	0.990	0.995	1.0	0.982
Tungiasis	0.989	1.0	0.995	0.995	1.0	1.0	1.0	1.0	1.0	1.0

[a]MLP: multilayer perceptron.
[b]NPV: negative predictive value.

Class Imbalance Handling

During the third experimental setting, demonstrating the class balancing method, 7 models (including NB) showed stable performance compared to the previous experiment with a missingness indicator. MLP, however, achieved improved results with perfect scores across all metrics, as shown in Table 7.

XSL•FO
RenderX

Table . Model performance results when trained using the *k*-fold cross-validation method.

Model	Accuracy		Macro average						
	Overall accuracy	Balanced accuracy	Precision	Recall	F_1-score	G-mean	ROC-AUC[a] (OVO[b])	ROC-AUC (OVR[c])	AUPRC[d]
Naïve Bayes	0.997	0.974	0.996	0.974	0.985	0.986	0.981	0.986	0.972
SVM[e] (linear)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
MLP[f]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Random forest	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
AdaBoost[g]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
CatBoost[h]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
XGBoost[i]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
LightGBM[j]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

[a]ROC-AUC: area under the receiver operating characteristic curve.
[b]OVO: one-vs-one.
[c]OVR: one-vs-rest.
[d]AUPRC: area under the precision-recall curve.
[e]SVM: support vector machine.
[f]MLP: multilayer perceptron.
[g]AdaBoost: adaptive boosting.
[h]CatBoost: categorical boosting.
[i]XGBoost: extreme gradient boosting.
[j]LightGBM: light gradient boosting machine.

In terms of per-class scores, all 7 models (except NB) scored 1.0 across all 5 per-class metrics, as shown in Table 8. NB showed lower per-class scores compared to the other 7 models.

Table . Model performance results when trained using the *k*-fold cross-validation method.

Class	SVM[a], RF[b], AdaBoost[c], CatBoost[d], XGBoost[e], LightGBM[f], MLP[g]					Naïve Bayes				
	Precision	Recall	F_1-score	Specificity	NPV[h]	Precision	Recall	F_1-score	Specificity	NPV
Podo	1.0	1.0	1.0	1.0	1.0	1.0	0.92	0.96	1.0	0.997
Scabies	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Tungiasis	1.0	1.0	1.0	1.0	1.0	0.989	1.0	0.995	0.995	1.0

[a]SVM: support vector machine.
[b]RF: random forest.
[c]AdaBoost: adaptive boosting.
[d]CatBoost: categorical boosting.
[e]XGBoost: extreme gradient boosting.
[f]LightGBM: light gradient boosting machine.
[g]MLP: multilayer perceptron.
[h]NPV: negative predictive value.

The Harmonized Approach

In this last experimental setting, we applied our proposed harmonized approach that integrates data balancing and robust validation methods, which presents 2 separate evaluation results: first, the performance scored on the test set, and second, the results scored on the nested CV method that incorporates the hyperparameter tuning process. The results in Table 9 present the overall test performance of the models of the final experiment using the hybrid approach. As shown in Table 9, NB and MLP scored similar performance scores of 0.974 in balanced accuracy, with the macro average scores of 0.996, 0.974, 0.985, and 0.986 in precision, recall, F_1-score, and G-mean, respectively. MLP scored higher results in ROC-AUC

XSL•FO
RenderX

and AUPRC, while the remaining 6 models scored perfect scores (1.0) across all evaluation metrics.

Table . Model performance results when trained using the *k*-fold cross-validation method.

Model	Accuracy		Macro average						
	Overall accuracy	Balanced accuracy	Precision	Recall	F_1-score	G-mean	ROC-AUC[a] (OVO[b])	ROC-AUC (OVR[c])	AUPRC[d]
Naïve Bayes	0.997	0.974	0.996	0.974	0.985	0.986	0.981	0.986	0.972
SVM[e] (linear)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
MLP[f]	0.997	0.974	0.996	0.974	0.985	0.986	0.997	1.0	0.999
Random forest	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
AdaBoost[g]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
CatBoost[h]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
XGBoost[i]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
LightGBM[j]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

[a]ROC-AUC: area under the receiver operating characteristic curve.
[b]OVO: one-vs-one.
[c]OVR: one-vs-rest.
[d]AUPRC: area under the precision-recall curve.
[e]SVM: support vector machine.
[f]MLP: multilayer perceptron.
[g]AdaBoost: adaptive boosting.
[h]CatBoost: categorical boosting.
[i]XGBoost: extreme gradient boosting.
[j]LightGBM: light gradient boosting machine.

Class-wise, 6 models (SVM, RF, AdaBoost, CatBoost, XGBoost, and LightGBM) achieved perfect scores (1.0) across all per-class metrics of precision, recall, F_1-score, specificity, and NPV, as presented in Table 10. Conversely, the MLP and NB models scored exactly the same per-class metrics, scoring the least sensitivity for podoconiosis with a disease-specific recall of 0.92, while scoring almost perfect scores in precision (0.99), specificity (0.995), and F_1-score (0.995) for the tungiasis class.

Table . Model performance results when trained using the *k*-fold cross-validation method.

Class	Models: SVM[a], RF[b], AdaBoost[c], CatBoost[d], XGBoost[e], and LightGBM[f]					Models: NB[g] and MLP[h]				
	Precision	Recall	F_1-score	Specificity	NPV[i]	Precision	Recall	F_1-score	Specificity	NPV
Podo	1.0	1.0	1.0	1.0	1.0	1.0	0.923	1.0	1.0	0.997
Scabies	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Tungiasis	1.0	1.0	1.0	1.0	1.0	0.989	1.0	0.995	0.995	1.0

[a]SVM: support vector machine.
[b]RF: random forest.
[c]AdaBoost: adaptive boosting.
[d]CatBoost: categorical boosting.
[e]XGBoost: extreme gradient boosting.
[f]LightGBM: light gradient boosting machine.
[g]NB: Naïve Bayes.
[h]MLP: multilayer perceptron.
[i]NPV: negative predictive value.

XSL•FO
RenderX

Nested CV and Hyperparameter Tuning

Nested CV, the method we applied for hyperparameter tuning and to achieve unbiased performance estimates, resulted in slight deviations compared to the test performance, as summarized in Table 11.

As shown, 5 models scored perfect scores across all metrics with no performance variability (σ=0) on the outer folds. However, 2 models (LightGBM and XGBoost) exhibited slight drops in their average scores of balanced accuracy and macro recall (Δ−0.007, σ=−0.013), and F_1-score (Δ 0.004, σ=−0.007), while the other scores remain stable.

Table . Summary of models' performance on the nested cross-validation method, with mean (SD) of key performance metrics.

Model	Balanced accuracy	Recall (macro)	F_1-score (macro)	Average precision	ROC-AUC[a] (OVO[b])
Naïve Bayes	0.999 (0.003)	0.999 (0.003)	0.999 (0.003)	0.998 (0.004)	0.999 (0.002)
SVM[c] (linear)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
MLP[d]	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
Random forest	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
AdaBoost[e]	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
CatBoost[f]	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
XGBoost[g]	0.993 (0.013)	0.993 (0.013)	0.996 (0.007)	1.0 (0.0)	1.0 (0.0)
LightGBM[h]	0.993 (0.013)	0.993 (0.013)	0.996 (0.008)	1.0 (0.0)	1.0 (0.0)

[a]ROC-AUC: area under the receiver operating characteristic curve.
[b]OVO: one-vs-one.
[c]SVM: support vector machine.
[d]MLP: multilayer perceptron.
[e]AdaBoost: adaptive boosting.
[f]CatBoost: categorical boosting.
[g]XGBoost: extreme gradient boosting.
[h]LightGBM: light gradient boosting machine.

Analysis of Results

Overall Performance and Model Stability

The second experiment that applied the simple imputation method for the structural missing values introduced the first performance variation, allowing the 4 models (NB, SVM, MLP, and AdaBoost) to have full performance scores. The third experiment brought the next variations, creating 3 divergent performance score groups after the application of disease-specific symptom applicability indicator flags. This experiment resulted in substantial performance changes, particularly to MLP, as evidenced by the model's absolute improvements of Δ+0.097 in balanced accuracy and macro recall, Δ+0.560 (σ=0.396) in podo precision, Δ+0.460 (σ=0.325) in podo F_1-score, Δ+0.295 (σ=0.209) in tungiasis precision, and Δ+0.173 (σ=0.122) in tungiasis F_1-score. According to the results, all improvements, particularly the class-specific scores, were achieved at a slight expense of the scabies class recall Δ−0.005 (σ=−0.004) and F_1-score Δ−0.002 (σ=−0.001). All these improved scores of MLP underscore that the feature engineering technique was significantly effective in helping the model identify the minority classes. The confusion matrices in Figures 2A and 2B depict all these changes shown by MLP across the 2 experiments. The other 6 models showed all perfect per-class scores with zero misclassification.

Unlike MLP, the generative model NB exhibited overall performance declines of Δ−0.026 in balanced accuracy and macro recall, Δ−0.015 in macro F_1-score, including precision (Δ−0.004) and G-mean (Δ−0.014), including the per-class scores. The remaining 6 models maintained the perfect scores across all metrics. The sample weighting method further improved the performance of MLP with an absolute increase of Δ+0.003 in balanced accuracy, recall, and G-mean, Δ+0.026 in macro F_1-score, and Δ+0.044 in precision, while the remaining 7 models remained consistent.

Overall, across all the experimental phases, 4 models (SVM, RF, CatBoost, and AdaBoost) showed absolute stability of performance, with no performance variation (σ=0.0), while the other 4 models (NB, MLP, XGBoost, and LightGBM) exhibited slight performance variations. Table 12 presents an analysis of models' performance variance across the 6 major evaluation metrics throughout the 5 experiments. As shown, MLP has the highest overall performance variance in its balanced accuracy and macro recall (σ=0.043), macro F_1-score and precision (σ=0.099), including G-mean (σ=0.03) and macro AUPRC (σ=0.001). Conversely, NB exhibited dropping patterns in balanced accuracy and macro recall (σ=−0.014), macro F_1-score (σ=−0.008), and precision, G-mean, and AUPRC (σ=−0.002, −0.007, and −0.015, respectively). These overall declines were introduced due to the introduction of symptom applicability indicator flags, which created a set of highly correlated features,

XSL•FO
RenderX

leading to incompatibility with the model's probabilistic assumption [39].

According to the results of the nested CV method (Table 12), LightGBM and XGBoost exceptionally exhibited marginal performance drops. These 2 models dropped their performance by Δ−0.007 (σ=−0.003) in balanced accuracy and macro recall, Δ−0.004 (σ=−0.002) in F_1-score and G-mean, while showing no change in precision and AUPRC.

XSL•FO
RenderX

Figure 2. Confusion matrices of MLP showing model performance after the second and third experiments: (A) confusion matrix of MLP after the second experiment with simple imputation and (B) confusion matrix of MLP after the third experiment with feature engineering. MLP: multilayer perceptron.

XSL•FO
RenderX

Table . Analysis of models' performance based on 6 evaluation metrics across 5 experiments.

Model	Balanced accuracy	Macro average				
		Precision	Recall	F_1-score	G-mean	AUPRC[a]
Naïve Bayes	0.014	0.002	0.014	0.008	0.007	0.015
SVM[b] (linear)	0.0	0.0	0.0	0.0	0.0	0.0
MLP[c]	0.043	0.099	0.043	0.099	0.03	0.001
Random forest	0.0	0.0	0.0	0.0	0.0	0.0
AdaBoost[d]	0.0	0.0	0.0	0.0	0.0	0.0
CatBoost[e]	0.0	0.0	0.0	0.0	0.0	0.0
XGBoost[f]	0.003	0.0	0.003	0.002	0.002	0.0
LightGBM[g]	0.003	0.0	0.003	0.002	0.002	0.0

[a]AUPRC: area under the precision-recall curve.
[b]SVM: support vector machine.
[c]MLP: multilayer perceptron.
[d]AdaBoost: adaptive boosting.
[e]CatBoost: categorical boosting.
[f]XGBoost: extreme gradient boosting.
[g]LightGBM: light gradient boosting machine.

Similar patterns were seen in the class-wise performance of these 4 models, as shown in Table 13.

MLP has the highest per-class performance variance with σ=0.33, σ=0.26, σ=0.49, and σ=0.39 in podo-precision, F_1-score, specificity, and recall, respectively. These higher variations are caused by the lowest per-class scores achieved during the third experimental training that demonstrates the structural missing values using missingness indicator features. NB exhibited smaller variance in tungiasis precision (σ=−0.006), podo recall (σ=−0.04), podo-F_1-score (σ=−0.023), tungiasis-F_1-score, and specificity (σ=−0.003). However, the 6 models maintained a totally stable per-class performance (σ=0).

XSL•FO
RenderX

Table . Summary of models' performance on the nested cross-validation method, with mean (SD) of key performance metrics.

Model and disease class	Per-class metrics				
	Precision	Recall	F_1-score	Specificity	NPV[a]
SVM[b], RF[c], AdaBoost[d], CatBoost[e], XGBoost[f], LightGBM[g]					
Podo	0.0	0.0	0.0	0.0	0.0
Scabies	0.0	0.0	0.0	0.0	0.0
Tungiasis	0.0	0.0	0.0	0.0	0.0
Naïve Bayes					
Podo	0.0	0.040	0.023	0.0	0.002
Scabies	0.0	0.0	0.0	0.0	0.0
Tungiasis	0.006	0.0	0.003	0.003	0.0
MLP[h]					
Podo	0.329	0.039	0.255	0.049	0.002
Scabies	0.0	0.005	0.002	0.0	0.009
Tungiasis	0.006	0.147	0.086	0.003	0.061

[a]NPV: negative predictive value.
[b]SVM: support vector machine.
[c]RF: random forest.
[d]AdaBoost: adaptive boosting.
[e]CatBoost: categorical boosting.
[f]XGBoost: extreme gradient boosting.
[g]LightGBM: light gradient boosting machine.
[h]MLP: multilayer perceptron.

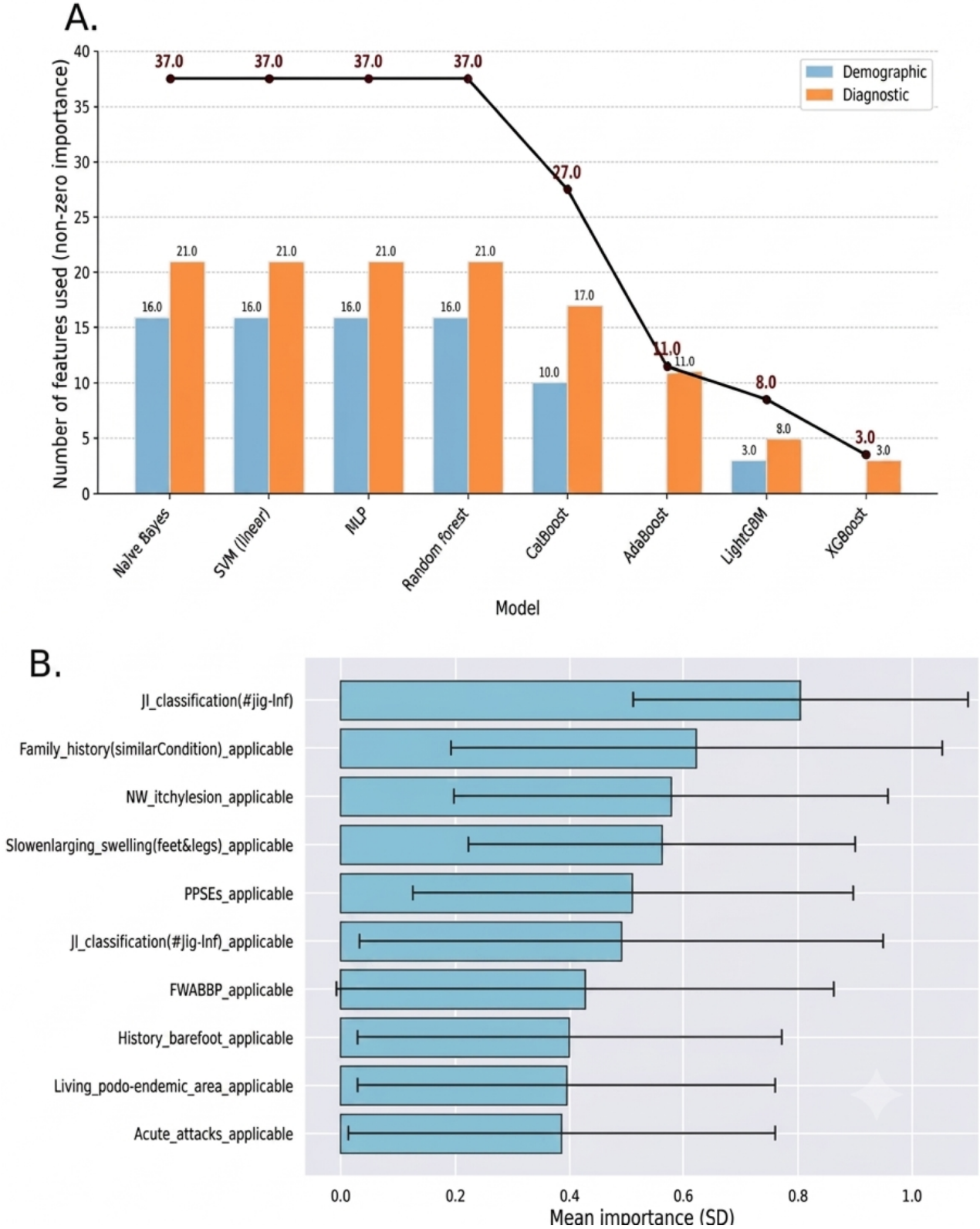
Analysis of Feature Importance

While the complete FI scores of the 38 features across 8 models are presented in Table S3 of Multimedia Appendix 2, in this section, we present an overall summary of the FI scores. Accordingly, the FI evaluation results revealed that 4 models (NB, SVM, MLP, and RF) used 37 features out of the 38 trainable features, as shown in Figure 3A.

Regarding individual features used, Figure 3B shows the top 10 features according to their use by the 8 models across all experiments. Accordingly, the jigger infestation classification feature (JI_classification(#jig-inf)) appears to be the top feature having the highest mean FI value. This feature appears in 6 models in their top 5 models list. Model-wise, we conducted deeper analyses of the feature relevance evaluation scores, presented in Table 14, that reveal a diverse set of feature distribution, utilization, and selectivity patterns among models.

XSL•FO
RenderX

Figure 3. Feature importance evaluation results of the MLP model: (A) overall analysis of the number of features used by each model and (B) top features based on the average importance scores across models. MLP: multilayer perceptron.



As shown, all the 5 tree-based models exhibited a right-skewed FI distribution, which signifies that smaller subsets of the entire feature set are highly relevant, with multiple remaining features being least or nonrelevant [40]. This demonstrated that these 5 tree-based models are highly selective in using features, as evidenced by the feature preference and sparsity ratio scores. Conversely, NB and MLP exhibited a left-skewed FI distribution pattern, highlighting that these models equally designated the majority of the features as highly relevant—MLP assigned the same FI score of 0.665 for 27 features, whereas NB assigned a uniform 1.0 FI score for 36 features. These models have higher nondiscriminative tendencies to identify and ignore less relevant features, as shown by the higher feature preference and lower sparsity ratio scores (Table 14). Finally, SVM is the only model

XSL•FO
RenderX

that demonstrated an approximately balanced (symmetric) FI score distribution, as shown by its lower sparsity and information decay scores. Overall, the model-specific FI distributions displayed 3 different distribution trends, which is confirmed by the aggregate summary of FI distribution results (mean 0.311, SD 0.390)—verifying that the overall distribution of the skin NTD diagnostic features is right-skewed. Given the higher mean (μ=0.311) having an SD that exceeds the mean FI score (σ=0.390), which creates higher aggregate variance factors, the right-skewed distribution is naturally expected to be reflected [40].

Table . Summary of feature importance scores and overall feature usage analysis of the 8 models.

Model	Diagnostic features (n=21)	Demographic features (n=17)	Distribution of FI[a] score	FI distribution type	Feature preference	Sparsity ratio	Information decay	Information decay type
Naïve Bayes	21	16	Mean 0.974, SD 0.160, range 0-1	Left-skewed	0.9737	0.0263	−0.004	Slow decay
SVM[b] (linear)	21	16	Mean 0.310, SD 0.334, range 0-1	Approximately symmetric	0.31	0.0526	−0.0284	Slow decay
MLP[c]	21	16	Mean 0.660, SD 0.160, range 0-1	Left-skewed	0.6604	0.0263	−0.0085	Slow decay
Random forest	21	16	Mean 0.232, SD 0.292, range 0-1	Right-skewed	0.2321	0.2105	−0.0239	Slow decay
CatBoost[d]	17	10	Mean 0.069, SD 0.200, range 0-1	Right-skewed	0.0692	0.2895	−0.0096	Slow decay
AdaBoost[e]	11	0	Mean 0.105, SD 0.201, range 0-1	Right-skewed	0.1045	0.7105	−0.013	Slow decay
LightGBM[f]	5	3	Mean 0.061, SD 0.178, range 0-1	Right-skewed	0.061	0.7895	−0.0086	Slow decay
XGBoost[g]	3	0	Mean 0.079, SD 0.270, range 0-1	Right-skewed	0.0789	0.9211	−0.0115	Slow decay
Overall aggregated summary	21	17	Mean 0.311, SD 0.390, range 0-1	Right-skewed	0.311 (0.313)	0.378 (0.347)	−0.0134 (0.0078)	Slow decay

[a]FI: feature importance.
[b]SVM: support vector machine.
[c]MLP: multilayer perceptron.
[d]CatBoost: categorical boosting.
[e]AdaBoost: adaptive boosting.
[f]LightGBM: light gradient boosting machine.
[g]XGBoost: extreme gradient boosting.

Discussion

While prior studies proposing AI-assisted diagnosis for skin NTDs primarily focused on image-based approaches, the use of structured clinical skin NTD data, the methodological implications, and enhancement strategies remained to be the major gaps. This preliminary pilot study demonstrated ML-based diagnosis of skin NTDs based on structured patient data, marking a high contribution to the problem domain, through the findings and their implications highlighted hereunder.

Important Findings

Novel Dataset and Diversified Data-Related Challenges

In this study, we crafted a new structured skin NTD clinical dataset, confronted by multifaceted data-related challenges (sample size limitation, severe class imbalance, and lack of full disease representation). We used our novel dataset in 3 different structures with the aim of experimentally analyzing, comparing, and identifying the best-performing ML model for the diagnosis of skin NTDs based on tabular patient data. In achieving the objectives, the study recorded multiple findings with different implications. The use of the initial dataset (having the structural

XSL•FO
RenderX

missing values) clearly demonstrated the resiliency of the 4 tree-based models (RF, CatBoost, LightGBM, and XGBoost) to null values due to their built-in mechanisms of handling missing values, which include considering the missing values as a separate group [41,42]. These results underlined the effectiveness of these tree-based architectures to establish robust real-time diagnostic tools, including the scenarios of having incomplete diagnostic data.

Methodological Rigor

Given the size and distribution of the new dataset, our study underlined that the selection and utilization of sound ML methods is the key to beat expected generalizability challenges, as demonstrated by the progressive performance gains across experiments. As an initial experimental finding, the Δ+0.097 (10%) and Δ+0.204 (24%) macro recall and F_1-score boosts of MLP validate the adoption of feature engineering as one big initial strategy in addressing and demonstrating our methodological success. Another important finding highlighted that some models, specifically LightGBM and XGBoost, exhibited unreliable performance due to predictive bias, a situation where the standard validation methods (the test CV performance in our case) were pretentiously presenting biased prediction estimates [38,39]. Such performance traps were exposed by the nested CV scores, proving the robustness of the dual CV-based approach in establishing unbiased disease prediction estimates, given the highly constrained dataset used in this study.

Overall, this study highlighted that stable model performances with higher predictive accuracies are achievable through a synergistic integration of multiple ML strategies working in harmony, instead of individual and isolated methods—first, establishing a reliable data preprocessing pipeline followed by a robust validation pipeline incorporating a synchronized implementation of class, all applied on top of the robust data preprocessing pipeline. However, as all these results are achieved against all the odds of data-related constraints, it is mandatory to have further investigations with extended resources to maintain the high achievements.

Overall Performance of Models

Based on the overall test performance scores, the majority of the models (except NB and MLP) exhibited consistent performance scores across all metrics. While desirable, perfect scores are not always good signs, given the dataset constraints of this study. Specifically, the perfect scores achieved by the highly feature-selective tree-based models, such as XGBoost, that used only 3 features for disease prediction, raise critical performance concerns. Positively, the perfect scores indicate the maximum disease discriminative power of the model. However, given the fact that the model makes disease predictions based on only 3 features out of 38, and these scores showed declines during the nested CV training, these perfect scores have negative implications, indicating that these scores are fragile and misleading. Such scores indicate that the models are facing the problem of over-optimism in their overall test performance. The use of such models for the diagnosis of diseases using only 3 diagnostic features (3 different symptoms of 3 different diseases) is highly risky and clinically infeasible, as the model would fail if one symptom is missing or misrepresented.

Implications of Results for Model Selection and Clinical Practices

Performance vs Characteristics of Models

According to the FI analysis reports, the diagnostic features have the highest importance for the classification of the skin NTDs and are most widely used by all models, while the demographic features are the least frequently used set of features for the classification of the skin NTDs. These results highlight that models using almost the whole feature set, such as random forest, achieved 100% performance scores across all metrics, while LightGBM (8 features) and XGBoost using only 3 features scored the same 100% performance. These tendencies of the models in using the least proportions of the trainable features could suggest that the models are applying complex architectures and are suffering from overfitting, given the limited dimensionality of our dataset used. Conversely, the tendencies of models to use the highest proportion of the trainable features could also point to the same problem, though such challenges can be alleviated through the use of ML-based strategies, as in the case of this study. Therefore, all these factors suggest that tradeoffs need to be made for the final model selection. Overall performance analysis and FI scores revealed that the 8 models exhibited different behaviors, forming 3 groups of different model characteristics.

All-Inclusiveness

Four models (NB, SVM, MLP, and RF) generally fall under this group due to the behavior they commonly share by equally ranking 37 features, while each model has specific determining characteristics. Accordingly, NB showed a different (outlier) behavior due to its invariable selection character that considers all features are equally important, having no interdependency or highly correlated relationship. This high-level feature-inseparability raises an issue, as it is not feasible for 2 or more features to be equally important in diagnosing skin NTDs. For instance, according to the results of our experimental results, a given value of the “educational level” of all patients has never been equally important as the diagnostic features, as confirmed by the dominant diagnostic feature “JI_classification(#Jig-Inf)” and its applicability flag. Further, the introduction of the symptom applicability flags created performance distortions. Both these factors diluted the reliability of the NB models, making them unable to differentiate between symptoms. This, in turn, results in a model inability to differentiate between diseases, which led to perfect scores, where such models cannot be reliable candidates for clinical diagnostics. MLP is the other all-inclusive model with a slightly different feature ranking behavior compared to NB, where 27 features (14 diagnostic and 13 demographic) showed an equal feature rank of 0.665. This depicts invariable ranks (σ=0) among the 27 features, highlighting that the model failed to effectively differentiate disease signal from noise, which also proved that this neural network-based model showed overfitting due to the model’s complex representational ability, given our small-sized tabular dataset. The other 2 models (SVM and RF) showed similar patterns of feature ranking scores with increasing feature

XSL•FO
RenderX

sparsity, where the variance in FI scores increased among features.

Shortcut Learners

Results showed that 3 of the tree-based models (XGBoost, LightGBM, and AdaBoost) exhibited a "lazy learning" pattern by using a limited number of features (3, 8, and 11 features), while all these 3 models achieved perfect scores. Especially with XGBoost showing absolute feature parsimony (showing high feature selectivity), the perfect scores using only 3 diagnostic features highlight a potential tendency of following shortcut paths by memorizing the feature patterns with the highest performance. Given the dataset constraints and the complexity of these models, this can underscore the plausibility of this shortcut-based (lazy) learning tendency by these 3 models, as exceptionally exposed by the overall performance declines recorded during the nested CV experiment.

Final Model Selection: Optimal Classifier

After the extensive analysis of the models' performance, the study concludes by identifying the model that achieved the optimal performance as the final skin NTD diagnostic model based on structured clinical data. According to the analytic results of the feature preference scores, CatBoost demonstrated the average performance standard compared to all models in our study. Statistically, based on the number of features identified as important (27 features, 71%), the distribution of the features (17 diagnostic and 10 demographic), and performance stability across experimental phases including nested CV, CatBoost showed an optimal separation pattern, unlike the other 2 groups of models. Additionally, its overall internal working logic of ordered boosting (to reduce prediction variances) and inherent categorical feature handling [40], including its fast and stable inference capabilities [43], could also play vital roles in helping the model achieve higher and more stable performance based on an optimal set of features. Therefore, we identified and proposed the CatBoost model as the potential benchmark skin NTD diagnostic model.

Conclusions

This study developed an ML-based diagnostic model for skin NTDs using a new tabular skin NTD diagnostic dataset (IDS). While the use of FDS (dataset created by applying only data preprocessing) resulted in perfect scores in all models except MLP, the third dataset architecture (created by applying feature engineering on FDS) created a new challenge for NB (due to statistical assumptions of the model), while stabilizing the performance of the other models. These data are used to experimentally analyze, compare, and identify the dataset structure with an optimal set of features, handle structural missing data, control overfitting, and achieve optimal model performance. The class weighting method we used helped in controlling the impact of the severe class imbalance by stabilizing the models' performance. Though all 6 models scored perfect scores across all metrics (except NB and MLP), the dual CV method proved its robustness, as the nested CV spotted hidden weaknesses in 2 of the tree-based models, XGBoost and LightGBM, showing slight performance declines while these models showed 100% test scores. These two models exhibited a high level of feature parsimony, while NB, SVM (Linear), MLP, and RF were highly feature inclusive. However, CatBoost showed optimal feature ranking and usage patterns. Overall, the boosting tree-based model CatBoost exhibited consistently higher performance, demonstrating optimal feature ranking and usage behavior, which substantiates the robustness of the model. Hence, we recommend further studies with extended resources using the CatBoost model, including RF and SVM as secondary alternatives, to demonstrate reliable diagnostic performance for further deployment.

While this study successfully established a robust ML pipeline as a methodological framework and developed a robust benchmark diagnostic model for skin NTDs, several inherent constraints created restraints to the study, limiting its contribution to its full potential. Data scarcity was the primary limitation of this study, while a severe class imbalance was also a constraint. Other limitations of the study include limited disease representation, inclusion of only one data modality, and specific geographic representation. Therefore, we recommend that further studies be conducted by collecting more patient data from multiple affected and potential areas, with the data being representative of all disease classes that are endemic to Ethiopia. This study also suggests the use of DL-based methods and the inclusion of other patient data (such as images and laboratory results), through proper balancing between diagnostic accuracy and computational expenses, to transform the diagnosis of skin NTDs into higher-level technology-assisted platforms and deliver quality health care services for affected areas, especially for resource-limited areas.

Acknowledgments

This study uses a novel dataset created using the data collected from patients with skin neglected tropical diseases (NTDs) living in Gacho Baba District of the Gamo Zone, Southwest Ethiopia. The data were initially collected for a project-based research intended to assess the burden of skin NTDs through community screening, led by Mr Alemayehu Bekele (Collaborative Research and Training Center for Neglected Tropical Diseases, Arba Minch University Medical College). We acquired the data for this study through institutional collaboration after having the required ethical clearance letter.

Therefore, the authors of this study gratefully acknowledge Mr Alemayehu Bekele and his team for providing the data to be used in this study, including the technical support he provided us.

All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided APF support for the

XSL•FO
RenderX

publication of this article. Finally, we (the authors this study) would like to declare that no generative tools and generated contents are used for the preparation of any part of this study.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of disease classes and missing values in the dataset.
[DOCX File, 119 KB - ojphi_v18i1e84966_app1.docx]

Multimedia Appendix 2
Tables presenting statistical analysis, feature importance evaluation scores, and descriptions.
[DOCX File, 68 KB - ojphi_v18i1e84966_app2.docx]

References

1. Elimination of neglected tropical diseases (NTDs) in Ethiopia – WOREDA level coordination toolkit for the WASH and NTD sectors. : Ethiopian Federal Ministry of Health; 2019 URL: https://www.susana.org/knowledge-hub/resources?id=3709 [accessed 2026-08-01]
2. Neglected tropical diseases. World Health Organization. 2025 Dec 14. URL: https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases [accessed 2026-08-01]
3. Ending the neglect to attain the sustainable development goals: a road map for neglected tropical diseases 2021–2030. : World Health Organization; 2020 Jan 28 URL: https://www.who.int/publications/i/item/9789240010352 [accessed 2026-08-01]
4. Abdela SG, Diro E, Zewdu FT, et al. Looking for NTDs in the skin; an entry door for offering patient centered holistic care. J Infect Dev Ctries 2020 Jun 29;14(6.1):16S-21S. [doi: 10.3855/jidc.11707] [Medline: 32614791]
5. Deribe K, Meribo K, Gebre T, et al. The burden of neglected tropical diseases in Ethiopia, and opportunities for integrated control and elimination. Parasit Vectors 2012 Oct 24;5(1):240. [doi: 10.1186/1756-3305-5-240] [Medline: 23095679]
6. Carrion C, Robles N, Sola-Morales O, Aymerich M, Ruiz Postigo JA. Mobile health strategies to tackle skin neglected tropical diseases with recommendations from innovative experiences: systematic review. JMIR Mhealth Uhealth 2020 Dec 31;8(12):e22478. [doi: 10.2196/22478] [Medline: 33382382]
7. The third national neglected tropical diseases strategic plan 2021-2025. : Ethiopian Federal Ministry of Health; 2021 Nov URL: https://espen.afro.who.int/sites/default/files/content/document/Third%20NTD%20national%20Strategic%20Plan%202021-2025_0.pdf [accessed 2026-08-01]
8. Marks M, Vedithi SC, van de Sande WWJ, et al. A pathway for skin NTD diagnostic development. PLoS Negl Trop Dis 2024 Nov;18(11):e0012661. [doi: 10.1371/journal.pntd.0012661] [Medline: 39585842]
9. Tadesse D, van Henten S, Batire S, et al. Decentralizing care for cutaneous leishmaniasis and other skin diseases to primary health facilities in Southern Ethiopia: what are the needs? BMC Infect Dis 2025;26:206. [doi: 10.1186/s12879-025-12324-0]
10. Yotsu RR, Almamy D, Vagamon B, et al. An mHealth app (eSkinHealth) for detecting and managing skin diseases in resource-limited settings: mixed methods pilot study. JMIR Dermatol 2023 Jun 14;6:e46295. [doi: 10.2196/46295] [Medline: 37632977]
11. Barbieri RR, Xu Y, Setian L, et al. Reimagining leprosy elimination with AI analysis of a combination of skin lesion images with demographic and clinical data. Lancet Reg Health Am 2022 May;9:100192. [doi: 10.1016/j.lana.2022.100192] [Medline: 36776278]
12. Achary R, Shelke CJ, Lekhya A. A GAN-enhanced multimodal diagnostic framework utilizing an ensemble of BiLSTM, BiGRU, and RNN models for malaria and dengue detection. Procedia Comput Sci 2025;252:381-393. [doi: 10.1016/j.procs.2024.12.039]
13. Yotsu RR. Integrated management of skin NTDs—lessons learned from existing practice and field research. Trop Med Infect Dis 2018 Nov 14;3(4):120. [doi: 10.3390/tropicalmed3040120] [Medline: 30441754]
14. Kokol P, Kokol M, Zagoranski S. Machine learning on small size samples: a synthetic knowledge synthesis. Sci Prog 2022;105(1). [doi: 10.1177/00368504211029777] [Medline: 35220816]
15. Xu P, Ji X, Li M, Lu W. Small data machine learning in materials science. npj Comput Mater 2023;9(1):1-15. [doi: 10.1038/s41524-023-01000-z]
16. Shwartz-Ziv R, Armon A. Tabular data: Deep learning is not all you need. Inf Fusion 2022 May;81:84-90. [doi: 10.1016/j.inffus.2021.11.011]

XSL•FO
RenderX

17. Florek P, Zagdański A. Benchmarking state-of-the-art gradient boosting algorithms for classification. arXiv. Preprint posted online on May 26, 2023. [doi: 10.48550/arXiv.2305.17094]
18. Wen Z, Li Q, He B, Cui B. Challenges and opportunities of building fast GBDT systems. In: Proceedings of the 30th International Joint Conference on Artificial Intelligence: IJCAI; 2021:4661-4668. [doi: 10.24963/ijcai.2021/632]
19. Wang X, Jia W. Optimizing edge AI: a comprehensive survey on data, model, and system strategies. arXiv. Preprint posted online on Jan 4, 2025. [doi: 10.48550/arXiv.2501.03265]
20. Aquil A, Saeed F, Baowidan S, Ali AM, Elmitwally NS. Early detection of skin diseases across diverse skin tones using hybrid machine learning and deep learning models. Information 2025 Feb;16(2):152. [doi: 10.3390/info16020152]
21. Jeong HK, Park C, Henao R, Kheterpal M. Deep learning in dermatology: a systematic review of current approaches, outcomes, and limitations. JID Innov 2023 Jan;3(1):100150. [doi: 10.1016/j.xjidi.2022.100150] [Medline: 36655135]
22. Silvey S, Liu J. Sample size requirements for popular classification algorithms in tabular clinical data: empirical study. J Med Internet Res 2024 Dec 17;26:e60231. [doi: 10.2196/60231] [Medline: 39689306]
23. Namburu A, Selvaraj P, Varsha M. Product pricing solutions using hybrid machine learning algorithm. Innov Syst Softw Eng 2022 Jul 25:1-12. [doi: 10.1007/s11334-022-00465-3] [Medline: 35910813]
24. Pattnayak P, Mohanty A, Das T, Patnaik S. Applying artificial intelligence and deep learning to identify neglected tropical skin disorders. In: 2024 3rd International Conference for Innovation in Technology (INOCON): IEEE; 2024. [doi: 10.1109/INOCON60754.2024.10511323]
25. Steyve N, Steve P, Ghislain M, Ndjakomo S, pierre E. Optimized real-time diagnosis of neglected tropical diseases by automatic recognition of skin lesions. Inform Med Unlocked 2022;33:101078. [doi: 10.1016/j.imu.2022.101078]
26. Recognizing neglected tropical diseases through changes on the skin: a training guide for front-line health workers. : World Health Organization; 2018 Jun 1 URL: https://www.who.int/publications/i/item/9789241513531 [accessed 2026-08-01]
27. Jackson J, Hofert M, Andrinopoulou E, et al. The challenge of handling structured missingness in integrated data sources. Adv Intell Discov 2025:e202500089. [doi: 10.1002/aidi.202500089]
28. Newman DA. Missing data: five practical guidelines. Organ Res Methods 2014 Oct;17(4):372-411. [doi: 10.1177/1094428114548590]
29. de Amorim LBV, Cavalcanti GDC, Cruz RMO. The choice of scaling technique matters for classification performance. Appl Soft Comput 2023 Jan;133:109924. [doi: 10.1016/j.asoc.2022.109924]
30. Metadata routing (“sample_weight”). scikit-learn. URL: https://scikit-learn.org/stable/metadata_routing.html [accessed 2026-08-01]
31. Waljee AK, Mukherjee A, Singal AG, et al. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 2013 Aug 1;3(8):e002847. [doi: 10.1136/bmjopen-2013-002847] [Medline: 23906948]
32. Zhuang N, Howells J. A computational pipeline for stratifying autoimmune patients using binary antibody data. bioRxiv. Preprint posted online on Dec 19, 2026. [doi: 10.1101/2025.09.11.670596]
33. Lumumba V, Kiprotich D, Mpaine M, Makena N, Kavita M. Comparative analysis of cross-validation techniques: LOOCV, k-folds cross-validation, and repeated k-folds cross-validation in machine learning models. Am J Theor Appl Stat 2024 Oct;13(5):127-137. [doi: 10.11648/j.ajtas.20241305.13]
34. Salmi M, Atif D, Oliva D, Abraham A, Ventura S. Handling imbalanced medical datasets: review of a decade of research. Artif Intell Rev 2024;57(10):273. [doi: 10.1007/s10462-024-10884-2]
35. Hr S, B A. Exploratory analysis of methods, techniques, and metrics to handle class imbalance problem. Procedia Comput Sci 2024;235:863-877. [doi: 10.1016/j.procs.2024.04.082]
36. Chen W, Yang K, Yu Z, Shi Y, Chen CLP. A survey on imbalanced learning: latest research, applications and future directions. Artif Intell Rev 2024;57(6). [doi: 10.1007/s10462-024-10759-6]
37. Palanivinayagam A, Damaševičius R. Effective handling of missing values in datasets for classification using machine learning methods. Information 2023 Feb;14(2):92. [doi: 10.3390/info14020092]
38. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? In: Proceedings of the 36th International Conference on Neural Information Processing Systems: Curran Associates; 2022. [doi: 10.52202/068431-0037]
39. Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. PeerJ Comput Sci 2016 Apr;2:e55. [doi: 10.7717/peerj-cs.55]
40. Lötsch J, Ultsch A. Recursive computed ABC (cABC) analysis as a precise method for reducing machine learning based feature sets to their minimum informative size. Sci Rep 2023 Apr 4;13(1):5470. [doi: 10.1038/s41598-023-32396-9] [Medline: 37016033]
41. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLOS ONE 2019;14(11):e0224365. [doi: 10.1371/journal.pone.0224365] [Medline: 31697686]
42. Parvandeh S, Yeh HW, Paulus MP, McKinney BA. Consensus features nested cross-validation. Bioinformatics 2020 May 1;36(10):3093-3098. [doi: 10.1093/bioinformatics/btaa046] [Medline: 31985777]
43. Filom K, Miroshnikov A, Kotsiopoulos K, Kannan AR. On marginal feature attributions of tree-based models. Found Data Sci 2024;6(4):395-467. [doi: 10.3934/fods.2024021]

XSL•FO
RenderX

Abbreviations

AUPRC: area under the precision-recall curve
CL: cutaneous leishmaniasis
CNN: convolutional neural network
CV: cross-validation
DL: deep learning
DTAG: Diagnostic Technical Advisory Group
FDS: final dataset
FEFDS: feature engineering on final dataset
FI: feature importance
IDS: initial dataset
KNN: *k*-nearest neighbors
LightGBM: light gradient boosting model
MAR: missing at random
MCAR: missing completely at random
MDA: mass drug administration
MICE: multivariate imputation by chained equation
ML: machine learning
MLP: multilayer perceptron
NB: naïve Bayes
NPV: negative predictive value
NTD: neglected tropical disease
PCA: principal component analysis
RF: random forest
SVM: support vector machine
TNR: true negative rate
TPR: true positive rate
WHO: World Health Organization
XGBoost: extreme gradient boosting

Edited by E Mensah; submitted 28.Sep.2025; peer-reviewed by A Famotire, KSM Tozammel Hossain; revised version received 25.Jun.2026; accepted 18.Jul.2026; published 21.Aug.2026.

Please cite as:
Minyilu Y, Yimer MA, Meshesha M
Automating Diagnosis of Skin Neglected Tropical Diseases via Patient Metadata through Machine Learning Model with Adaptive Balancing and Dual Cross-Validation: Retrospective Diagnostic Accuracy Study
Online J Public Health Inform 2026;18:e84966
URL: https://ojphi.jmir.org/2026/1/e84966
doi:10.2196/84966

© Yohannes Minyilu, Mohammed Abebe Yimer, Million Meshesha. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 21.Aug.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

eHealth Approaches to Multimorbidity Management: Scoping Review of Interventions and Multidisease Strategies

Katja Bochtler[1,2], Prof Dr; Salima Houta[2,3], MSc

[1]Faculty of Computer Science, Kempten University of Applied Sciences, Bahnhofstr. 61, Kempten, Bavaria, Germany

[2]Bavarian Center for Digital Health and Social Care, Kempten, Bavaria, Germany

[3]Department of Computer Science, TU Dortmund University, Dortmund, North Rhine-Westphalia, Germany

Corresponding Author:
Katja Bochtler, Prof Dr
Faculty of Computer Science, Kempten University of Applied Sciences, Bahnhofstr. 61, Kempten, Bavaria, Germany

Abstract

Background: Multimorbidity, defined as the coexistence of two or more chronic conditions, challenges traditional, single-disease health care models. Although digital health technologies are increasingly applied to complex patient care, most interventions remain disease-specific and fail to account for interactions between coexisting conditions.

Objective: This scoping review aimed to systematically map and synthesize existing evidence on digital health interventions that explicitly address multimorbidity, focusing on how these solutions integrate disease interdependencies between coexisting diseases in their design, implementation, and evaluation. We aimed to identify underlying concepts, map frequently targeted disease clusters, and assess reported outcomes and barriers to real-world application.

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) framework and informed by PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) recommendations, we conducted systematic searches in PubMed and CINAHL for studies published in English or German between January 2014 and May 2024. Eligible studies described digital health interventions targeting patients with two or more conditions and explicitly addressed interactions among diseases, symptoms, or treatments. We strictly excluded studies of digital solutions that addressed multiple conditions through isolated modules or separate functionalities without cross-condition integration. In addition, research focusing solely on technical development, polypharmacy, or usability was excluded. Two reviewers independently screened records and extracted data using a structured charting form. Data were synthesized using descriptive statistics and thematic analysis.

Results: Of 1660 records identified, 11 studies representing seven distinct digital health projects met the inclusion criteria. Most interventions focused on older adults with cardiometabolic or respiratory multimorbidity and combined mobile apps, telemonitoring platforms, and decision-support or analytic modules. While most interventions demonstrated feasibility and user acceptance, only a minority operationalized disease interactions beyond aggregated data. Approaches to multimorbidity included cross-disease analytics, composite risk indicators, and guideline-based decision support; however, systematic reconciliation of competing clinical recommendations was rare. Evaluation outcomes mainly addressed usability, engagement, and feasibility, whereas evidence regarding clinical effectiveness remained limited and inconsistent. Only two studies assessed clinical outcomes, with mixed findings.

Conclusions: This review provides a novel synthesis of how digital health interventions explicitly address multimorbidity as an interconnected phenomenon rather than as a set of isolated conditions. Current evidence suggests feasibility and user acceptance of digital tools, whereas robust implementation of multimorbidity-aware functionalities and evidence regarding benefit remain limited. Advancing this field will require interoperable, data-integrative systems capable of supporting coordinated multimorbidity care across conditions and care settings. These findings highlight the need for a shift from disease-centric toward integrated multimorbidity-oriented digital care models.

(Online J Public Health Inform 2026;18:e88021) doi:10.2196/88021

KEYWORDS

multimorbidity; digital health; eHealth intervention; integrated care; multidisease approaches; disease interactions; self-management

XSL•FO
RenderX

Introduction

Background

Multimorbidity, commonly defined as the cooccurrence of two or more chronic conditions in an individual [1], represents one of the most significant and complex challenges facing contemporary public health systems. Its global prevalence has risen substantially due to population aging and lifestyle-related risk factors. Although multimorbidity is more frequent in older adults, more than 50% of individuals affected are under the age of 65 years [2]. A recent meta-analysis estimated the global prevalence of multimorbidity at 37.2%, with the highest rates in South America (45.7%), followed by North America (43.1%), Europe (39.2%), and Asia (35%) [3]. The impact of multimorbidity extends far beyond individual diagnoses, affecting physical function, quality of life, mental health, and mortality [2,4]. Managing multiple long-term conditions often involves complex and conflicting self-care regimens, frequent clinical appointments, and polypharmacy [5,6]. These interconnected health demands often surpass the capacity of current health care systems, which remain predominantly organized around single-disease paradigms [2,4,7,8]. There is increasing recognition of the need to transition from disease-centered to patient-centered care. Effective management of multimorbidity demands integrated, personalized, and coordinated care strategies that address the complex interplay of biological, psychological, and social dimensions of health [8,9].

Digital health technologies can support person-centered care. They offer scalable, flexible, and low-threshold access to care and information, particularly in resource-constrained environments. However, most existing solutions continue to focus on single diseases, often replicating the same disease-centric logic that dominates conventional health care [9,10]. While these applications may offer high usability and clinical value for specific conditions, they often overlook the multifaceted needs of patients with multimorbidities. Physicians have also expressed concern that many health apps and electronic health records remain fragmented and condition-specific, thereby impeding coordinated and holistic care [11,12], which directly contradicts European policy frameworks promoting integrated, cross-sectoral patient care and shared decision-making [13]. As a result, patients with multimorbidities often face the burden of navigating multiple disease-specific apps, platforms, and care plans.

Rationale

Previous reviews examined specific comorbidity pairs and digital health interventions across diseases, including communication tools, self-management support, telemedicine, and clinical decision support systems (CDSSs). Yet, evidence for truly integrated digital support in multimorbidity remains limited, as most studies addressed conditions in parallel rather than their interactions [14-22].

None of the studies examined how interventions operationalize disease interdependencies, such as modeling condition interactions or reconciling competing treatment recommendations, which is an essential feature of truly integrated multimorbidity care. Given the fragmented and heterogeneous nature of the literature, a scoping review was considered the most appropriate methodology.

Objectives

This scoping review synthesizes evidence on digital health solutions for multimorbidity management that explicitly address interactions between coexisting conditions. The primary and secondary objectives are given in Textbox 1.

Textbox 1. Primary and secondary objectives.

Primary objectives

- To identify and classify eHealth solutions for multimorbidity, focusing on cross-disease interactions and symptom interdependencies.
- To analyze underlying concepts, including strategies for integration, personalization, and care coordination.
- To map targeted diseases or multimorbidity clusters and identify recurring patterns.

Secondary objectives

- To assess reported effectiveness and implementation outcomes, including feasibility, usability, and clinical impact.
- To identify challenges and barriers to real-world scalability and sustainability.

Methods

Protocol and Registration

This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews [23]; see Checklist 1 for the completed report and abstract checklists). The reporting of the search strategy was additionally informed by the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) to enhance transparency and reproducibility. A formal review protocol was not prospectively registered.

Eligibility Criteria

Eligibility criteria were established a priori to ensure a systematic and targeted selection of relevant studies (Table 1). To ensure methodological rigor, the eligibility criteria were structured according to the Population, Concept, and Context (PCC) framework [24].

XSL•FO
RenderX

Table . Eligibility criteria using the Population, Concept, and Context framework.

Domain	Inclusion criteria	Exclusion criteria
Population	• Individuals with clinically diagnosed multimorbidity (≥2 chronic conditions) • Studies involving health care providers using multimorbidity cases (eg, vignettes and simulated profiles)	• Single-disease populations • Frailty-only populations without confirmed multimorbidity
Concept	• Digital health interventions for treatment, management, or monitoring of multimorbidity • Multicondition functionality addressing disease interactions	• No digital or eHealth component (eg, purely face-to-face care, paper-based interventions, telephone-only interventions without digital integration) • Single-disease interventions (even in multimorbid populations) • Technology used only for data collection and analysis • No explicit focus on disease interactions • Articles focusing solely on usability, usage behavior, polypharmacy, or cost analysis • Technical development studies without clinical, patient-centered, or system-level context
Context	• No restriction on application settings or geographic and cultural background	• No clear link to clinical multimorbidity management (eg, abstract models without application context for multimorbidity)
Types of evidence sources	• Peer-reviewed primary studies	• Editorials, commentaries, reviews, and protocols • Conference abstracts without full text • Unavailable full texts
Language	• English • German	• Other languages
Date range	• Studies published between January 1, 2014, and May 5, 2024	• Studies published outside date range

Information Sources

Systematic searches were conducted in PubMed and CINAHL for studies published between January 1, 2024, and May 5, 2024. These databases were selected because they provide comprehensive coverage of biomedical, clinical, and health services research, including nursing and care-related literature, which is particularly relevant for multimorbidity and integrated care contexts. Multidatabase platforms were not used for this search. Although the search was limited to two major databases, additional backward reference screening of included studies and relevant reviews was conducted to improve coverage. Searches were limited to English- and German-language publications. The final search was conducted on May 5, 2024.

Other methods, including the search of study registries, online resources, or handsearching of specific journals, were not part of the research methodology. No direct contact with authors was established.

Search

The search strategy was developed iteratively to balance sensitivity and specificity, informed by established guidance for evidence synthesis. Search terms combined concepts related to digital health (eg, eHealth, mobile health [mHealth], and telemedicine) and multimorbidity (eg, comorbidity and multiple chronic conditions).

To enhance completeness, a multistep approach was applied:

1. Identification of key publications through preliminary searches and prior domain knowledge.
2. Analysis of keywords and indexing terms from relevant studies to identify additional synonyms and related concepts.
3. Use of a general-purpose large language model (LLM; ChatGPT, OpenAI) during the exploratory development of the search strategy to generate additional candidate synonyms for search terms. The LLM was used exclusively to support terminology expansion; all generated terms were manually reviewed by the authors and only those deemed relevant were incorporated into the final search string.
4. Iterative testing and refinement of the search query in both databases with manual screening of results to optimize relevance.

The final search strings for each database are provided in Multimedia Appendix 1. Searches were conducted in titles, abstracts, and keywords in PubMed and in CINAHL in all text fields.

XSL•FO
RenderX

No prevalidated search filters were applied. This search was an original effort and did not build upon prior work or previous search updates. A formal peer review of the search strategy was not conducted.

Selection of Sources of Evidence

All references from the literature search were imported into a Citavi citation manager library (Lumivero). Deduplication was performed by comparing DOIs and titles. The screening process used the following steps:

1. Each reviewer independently screened all titles and abstracts against the defined eligibility criteria. A small subset of studies was reviewed jointly in advance to calibrate the reviewers and ensure a shared understanding and consistent application of the eligibility criteria, particularly regarding technologically supported disease interactions in multimorbidity management.
2. All articles deemed potentially relevant were retrieved in full and independently assessed by both reviewers.

Disagreements at both stages were resolved through discussion and consensus.

Data Charting Process

Data was extracted using a structured spreadsheet developed for this review. Two reviewers (KB and SH) independently charted the data and discussed uncertainties throughout the process to ensure consistent interpretation of the charting items. The charting form was refined iteratively as familiarity with the evidence increased.

Data Items

The extracted items were organized into five main domains: (1) general information (eg, authors, year, and country), (2) disease characteristics (eg, targeted diseases and degree of multimorbidity), (3) digital solutions (eg, type of technology and integration mechanisms), (4) evaluation methods (eg, study design and sample demographics), and (5) outcomes (eg, primary and secondary outcomes). A detailed list of all extracted data items is provided in Multimedia Appendix 2.

Synthesis of Results

Charted data were summarized descriptively using frequencies, ranges, and medians where appropriate. In addition, intervention characteristics were analyzed thematically and grouped into recurring categories relating to disease interactions, technological functionalities, care concepts, implementation approaches, and reported outcomes. This facilitated the identification of common patterns and differences across digital interventions for multimorbidity.

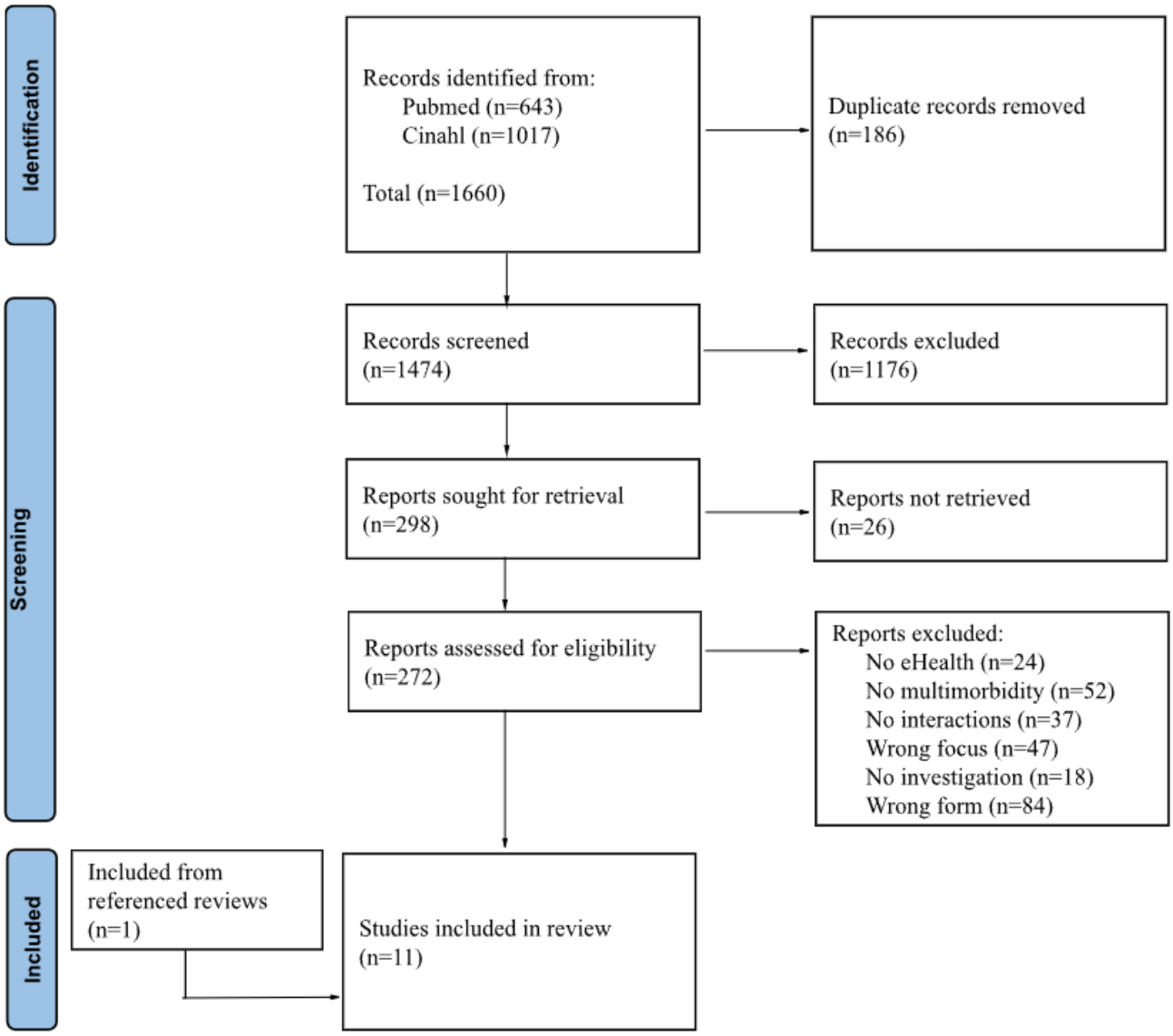
Results

Selection of Sources of Evidence

A total of 1660 studies were identified in our search across two databases: 643 (38.7%) from PubMed and 1017 (61.3%) from CINAHL. After removing duplicates, 1474 (88.8%) studies remained. Screening of abstracts and titles excluded 1176 (79.8%) studies that did not meet the inclusion criteria. Of the remaining 298 studies, 26 (1.57%) full-text papers could not be retrieved. Thus, 272 papers underwent a full-text review. Following this assessment, 10 (0.6%) studies met the inclusion criteria, and one relevant study was identified through screening of reference lists of review articles [25-35]. In total, 11 papers [25-35] were included in this review. The PRISMA diagram systemizing the screening process can be seen in Figure 1, as well as the reasons for full-paper exclusion of 262 (15.78%) studies.

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram of the study selection process for a scoping review of interventions and multidisease strategies accounting for disease interactions in multimorbidity (2014 - 2024).



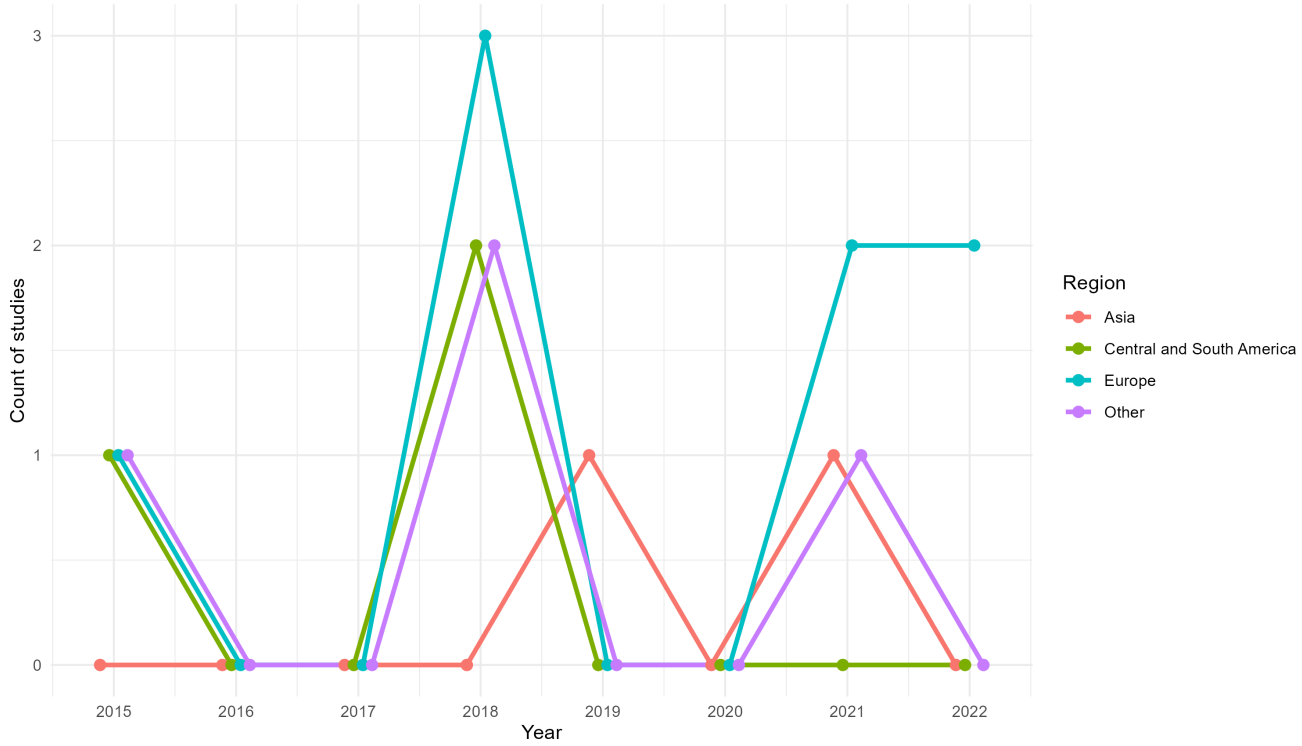
Characteristics of Sources of Evidence

In total, 11 studies [25-35] were included in this review. These are linked to seven unique digital health projects: Integrated Technology Systems for Proactive Patient-Centered Care (ProACT [25-27]), Mobile Airways Sentinel Network (MASK [28-30]), Multimorbidities Managing Technology for Healthcare (MATCH) [31], Autonomie trotz Multimorbidität in Sachsen durch Patientenempowerment, Holistische Versorgung für Ältere mit Vernetzung aller Regionalen Einrichtungen und Dienstleister (Autonomy despite Multimorbidity in Saxony through Patient Empowerment: Holistic Care for Older Adults via the Networking of All Regional Facilities and Service Providers; ATMoSPHÄRE [32]), Novel Model for Health Care Delivery (NOMHAD [33]), Mobile Atrial Fibrillation Application (mAFA [34]), and An integrated mHealth System for the Prevention and Care of Chronic Disease (mWellcare [35]).

The majority of studies included in this scoping review were conducted in European countries. Specifically, out of the 11 included studies, 9 (81.82%) [25-33] studies involved European countries, with common contributors being Ireland, Belgium, Germany, and France. Other represented regions included Asia (eg, China and India), Central and South America (eg, Mexico and Brazil), and Oceania (Australia; Figure 2).

XSL•FO
RenderX

Figure 2. Annual number of publications on digital support accounting for disease interactions in multimorbidity, by region (N=11; 2014 - 2024).



The number of included studies did not show a clear upward trend over time. While individual years such as 2018 and 2021 show a higher number of publications, the distribution across years remains relatively flat, suggesting that this research topic is not yet growing rapidly in the literature or is possibly underrepresented.

Results of Individual Sources of Evidence

Data extraction was conducted on 2 levels: study-specific information such as study design, sample size, and outcomes were documented per publication. In cases where multiple publications referred to the same project, data extraction was partially aggregated. The description of addressed diseases, target groups, technologies, functionalities, and data categories were conducted on a project level to capture the overall context and structure of the eHealth intervention. Study-specific and project-specific data extractions can be found in Multimedia Appendices 3 and 4.

Disease Characteristics

The seven projects included in the reviewed studies cover a range of chronic conditions. mAFA had the highest diversity of disease mentions (n=6), encompassing conditions such as atrial fibrillation (AF), coronary heart disease (CHD), congestive heart failure (CHF), depression, prior ischemic stroke, and pulmonary disease. Other projects such as ProACT (n=5) and mWellcare (n=4) also featured multiple diseases, indicating a broader scope in terms of comorbidity inclusion. In contrast, projects like MASK and MATCH featured the combination of two specific conditions. Although MASK also encompasses disease heterogeneity, it was considered eligible for this review because the intervention explicitly addresses the coexistence and interactions of asthma and allergic diseases within integrated multicondition care pathways rather than managing a single disease in isolation. The project ATMoSPHÄRE is not restricted to specific diseases or symptoms. The recommendations within this project are not medical in nature but rather focus on social services, including improving the overall condition and management of the disease, rather than recommending treatment.

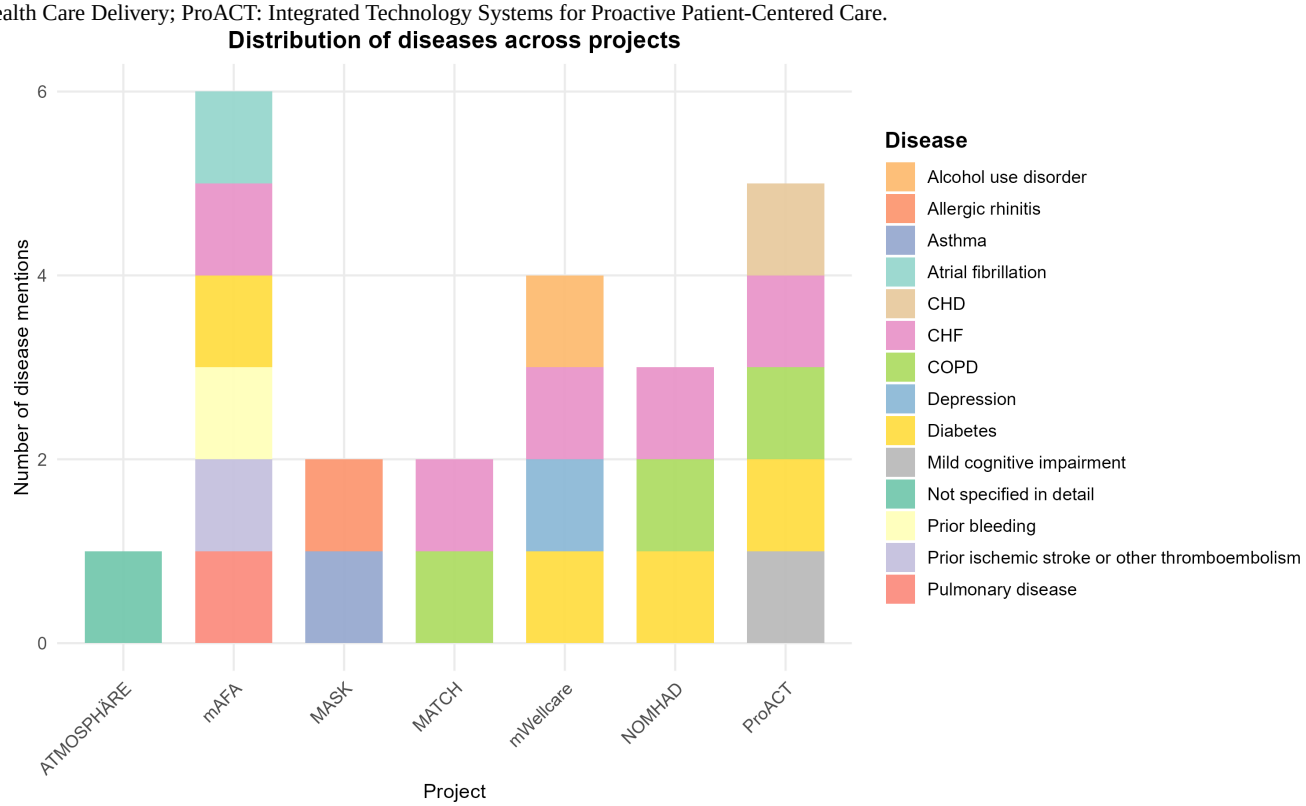
CHF (n=5) and diabetes (n=4) were the most frequently addressed diseases, often appearing together in studies focused on cardiovascular or metabolic multimorbidity. Chronic obstructive pulmonary disease (COPD; n=3) was the third most frequently mentioned condition.

The seven reviewed projects address varying levels of multimorbidity, reflecting different degrees of complexity in chronic disease management. Three projects (MASK, MATCH, and mWellcare) focused on specific combinations of two diseases, typically addressing well-defined comorbidity constellations such as cardiometabolic or respiratory disorders. These represent more straightforward cases of multimorbidity where clinical pathways and treatment strategies are relatively well established. In two projects (ProACT and NOMHAD), participation required the presence of two or more chronic conditions. For example, in ProACT, participants need to have at least two diseases out of a predefined disease set (CHD, CHF, COPD, diabetes, and mild cognitive impairment; Figure 3). This broader inclusion criterion reflects a more flexible design, capable of accommodating a wider range of patient needs while still maintaining a manageable level of system complexity. In one study (mAFA-Trial), inclusion required at least three chronic conditions from a predefined disease set, addressing a higher degree of multimorbidity and the corresponding need for more integrated care and decision support. Only one study (ATMoSPHÄRE) imposed no restrictions on the type of supported conditions, requiring only the presence of multimorbidity rather than specific disease combinations. Instead of providing disease-specific recommendations, it offered

XSL•FO
RenderX

supportive interventions based on comprehensive geriatric assessments.

Figure 3. Diseases targeted by digital health projects (n=7) accounting for multimorbidity interactions, derived from the included studies (N=11; 2014 - 2024) [25-35]. ATMoSPHÄRE: Autonomy despite multimorbidity in Saxony through patient empowerment: Holistic care for older adults via the networking of all regional facilities and service providers.; CHD: coronary heart disease; CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease; mAFA: Mobile Atrial Fibrillation Application; MASK: Mobile Airways Sentinel Network; MATCH: Multimorbidities Managing Technology for Healthcare; mWellcare: an integrated mHealth system for the prevention and care of chronic disease; NOMHAD: Novel Model for Health Care Delivery; ProACT: Integrated Technology Systems for Proactive Patient-Centered Care.



Classification of eHealth Solutions

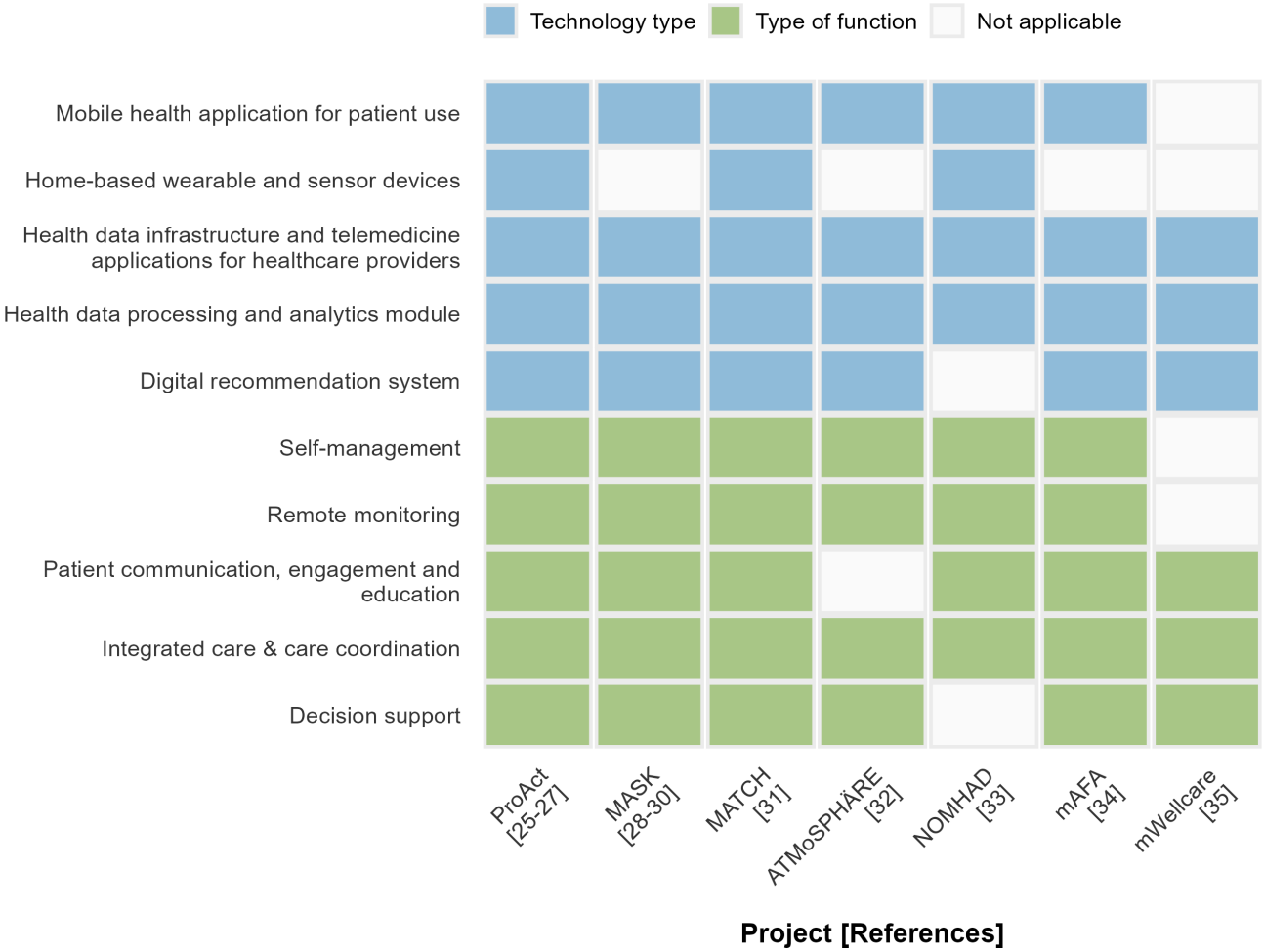
The classification of the eHealth interventions presented in the included studies was conducted from both a technical and functional perspective. For the categorization, we considered the World Health Organization (WHO) Classification of Digital Health Interventions v.1.0 as a foundational framework [36] but developed custom taxonomies to better capture the technological and functional aspects relevant to our analysis. In terms of technology type, we found five relevant system categories represented in the papers: (1) mHealth application for patient use, (2) home-based wearable and sensor devices, (3) health data infrastructure and telemedicine applications for health care providers, (4) health data processing and analytic modules, and (5) digital recommendation systems (DRSs).

In terms of which functions are supported by the technology, we found five categories: (1) self-management; (2) remote monitoring; (3) patient communication, engagement, and education; (4) integrated care and care coordination; and (5) decision support.

The most frequently reported technology categories were health data infrastructure and telemedicine applications for providers, followed by mHealth applications for patients and DRSs (Figure 4). Technologies related to mHealth for patients, clinical recommendation systems, and wearables were also commonly represented across the dataset. Notably, ProACT and MATCH reported the widest range of technology categories. In contrast, mWellcare focused on fewer technology categories, primarily emphasizing digital recommendation and telemedicine without involving patients through a mobile app. Regarding functions, integrated care and care coordination was the only functional category represented across all included projects, underlining its role as a fundamental component in managing patients with multimorbidities. The other functional categories were represented with nearly equal frequency across the dataset (Figure 4).

XSL•FO
RenderX

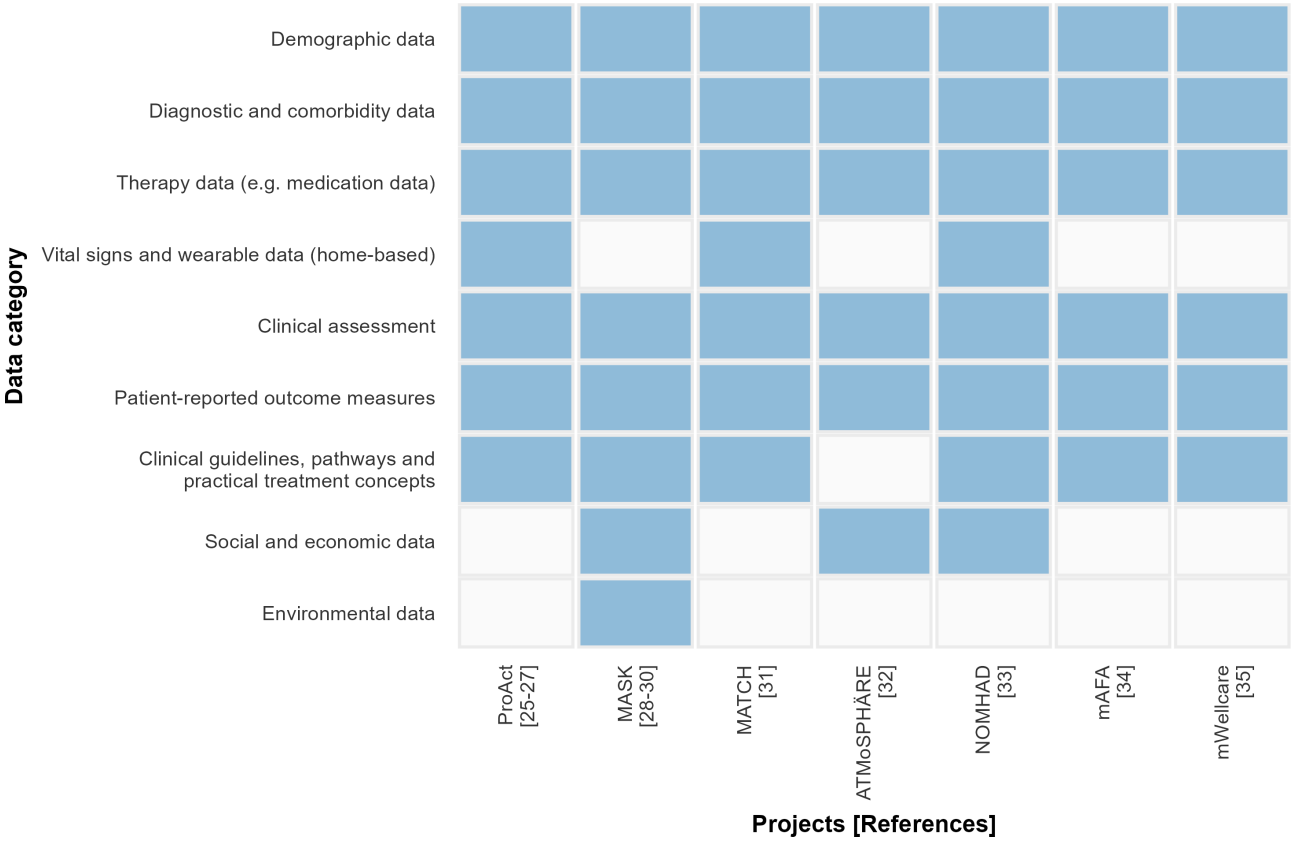
Figure 4. Characteristics of included projects (n=7): digital health technology types and functionalities for managing disease interactions reported in studies (N=11) on digital health solutions accounting for disease interactions in multimorbidity (2014 - 2024) [25-35]. ATMoSPHÄRE: Autonomy despite Multimorbidity in Saxony through Patient Empowerment: Holistic Care for Older Adults via the Networking of All Regional Facilities and Service Providers.; CHD: coronary heart disease; CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease; mAFA: Mobile Atrial Fibrillation Application; MASK: Mobile Airways Sentinel Network; MATCH: Multimorbidities Managing Technology for Healthcare; mWellcare: an integrated mHealth System for the Prevention and Care of Chronic Disease; NOMHAD: Novel Model for Health Care Delivery; ProACT: Integrated Technology Systems for Proactive Patient-Centered Care.



In addition to technologies and functionalities, the analysis identified the main data categories captured across the reviewed projects (Figure 5). The most commonly reported categories include demographic data, diagnostic and comorbidity data, therapy data, and clinical assessment. Following closely, vital signs and wearable data collected in home-based settings, patient-reported outcomes, as well as clinical guidelines along with practical treatment concepts were mentioned in the vast majority of projects. Some projects also consider social and economic factors (eg, social support network and work productivity), which are often relevant in the context of multimorbidity. MASK also integrates environmental data, such as pollen occurrence in regions, providing a holistic view of patient health within the living environment.

XSL•FO
RenderX

Figure 5. Characteristics of included projects (n=7): extracted data categories reported in studies (N=11) on digital health solutions accounting for disease interactions in multimorbidity management (2014 - 2024) [25-35]. ATMoSPHÄRE: Autonomie trotz Multimorbidität in Sachsen durch Patientenempowerment, Holistische Versorgung für Ältere mit Vernetzung aller Regionalen Einrichtungen und Dienstleister; CHD: coronary heart disease; CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease; mAFA: Mobile Atrial Fibrillation Application; MASK: Mobile Airways Sentinel Network; MATCH: Multimorbidities Managing Technology for Healthcare; mWellcare: An integrated mHealth System for the Prevention and Care of Chronic Disease; NOMHAD: Novel Model for Health Care Delivery; ProACT: Integrated Technology Systems for Proactive Patient-Centered Care.



Multimorbidity-Aware System Characteristics and Disease Interactions

The projects were classified based on their alignment with multimorbidity-aware system characteristics. While existing frameworks [37] highlight the need for structured analysis, they focus on clinical reasoning rather than technical system functionalities. We developed a classification scheme focusing on the functional depth of interventions. This categorization allows us to differentiate systems by their level of support, ranging from passive information display to active, prescriptive guidance. Aggregation of multimorbidity-related data is considered a prerequisite and therefore excluded from the classification.

1. Multimorbidity-aware analytics: Analytical methods designed to detect multimorbidity-related patterns, generate risk profiles for patients with multiple chronic conditions, and model potential interactions between diseases.
2. Multimorbidity-aware dashboard and alerts: Visualization and alerting tools tailored to represent the complexity of the state of a patient with multimorbidity. They generate context-sensitive alerts (eg, conflicts and risks) and highlight critical clinical issues (eg, overall risk indicator [ORI]) to assist health care professionals in making individual decisions.
3. Multimorbidity-aware digital recommendation: DRSs (eg, CDSSs) that integrate formal concepts (eg, multiple clinical guidelines), resolve conflicts between disease-specific recommendations, and generate individualized recommendations.

While the classification of multimorbidity-aware system characteristics defines how a system supports clinical decision-making, its practical impact is determined by what specific clinical complexities are being addressed. To provide a holistic view of the digital health interventions, we complement this classification with a clinical dimension centered on disease interaction.

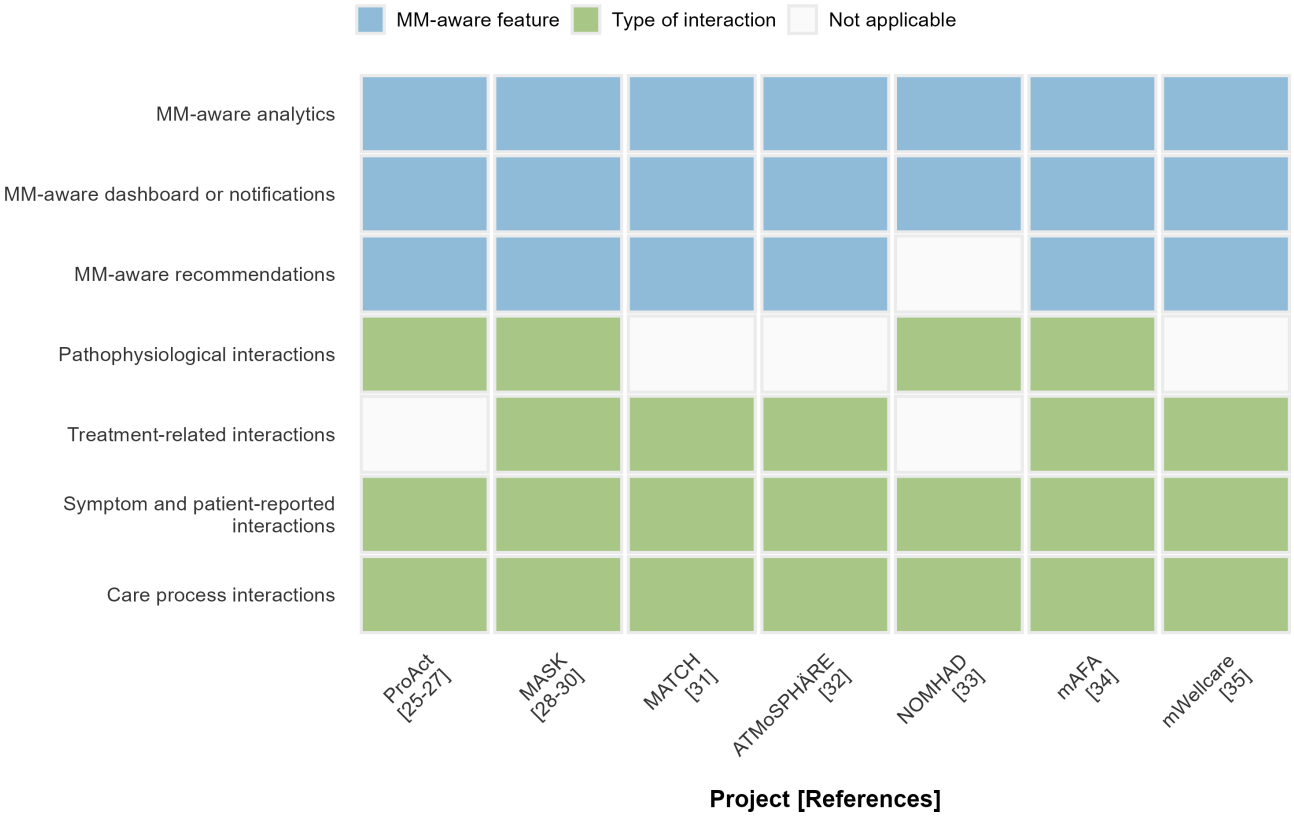
The conceptualization of “disease interactions” in multimorbidity is not uniformly defined in the literature but is discussed across several complementary perspectives. Prior work has highlighted biological interdependencies between coexisting conditions, challenges arising from competing treatment recommendations and polypharmacy, as well as the cumulative symptom burden experienced by patients and the need for coordinated, integrated care processes [5,6,8,11,13]. In addition, research on multimorbidity-oriented clinical decision support emphasizes the importance of integrating multiple data sources and care pathways when managing patients with complex chronic conditions [37].

XSL•FO
RenderX

Building on these perspectives, we developed an analytical framework to categorize how digital health interventions operationalize disease interactions. We distinguish four types of interactions: (1) pathophysiological interactions, referring to biological interdependencies between diseases; (2) treatment-related interactions, including conflicts or synergies between therapies or clinical guidelines; (3) symptom and patient-reported interactions, capturing overlapping or interacting symptoms and patient-reported outcomes; and (4) care process interactions, referring to coordination, prioritization, and integration of care activities across conditions.

The analysis of the included projects reveals varying degrees of multimorbidity awareness, ranging from basic data analysis to advanced decision support (Figure 6). Regarding multimorbidity-aware system characteristics, systems like NOMHAD and mWellcare focus on integrated data analysis and risk stratification, exemplified by NOMHAD's ORI. A higher level of functional depth is reported in ProACT, which uses inferential Bayesian Network (BN) modeling to identify latent biological variables. Similarly, MASK achieves depth by operationalizing the United Airway concept, using systematic pattern recognition to capture the physiological interplay between allergic conditions. Furthermore, projects like mAFA, MATCH, and ATMoSPHÄRE implement prescriptive decision support, using the ABC pathway (A: anticoagulation or avoid stroke; B: better symptom control; and C: cardiovascular disease and comorbidity management) or specialized decision support systems (DSSs) to provide structured clinical guidance. Unlike systems that actively suggest therapy adjustments, NOMHAD acts primarily as a risk-monitoring tool.

Figure 6. Characteristics of included projects (n=7): functional depth of multimorbidity-aware features and targeted disease interaction categories reported in studies (N=11) on digital health solutions accounting for disease interactions in multimorbidity (2014 - 2024) [25-35]. ATMoSPHÄRE: Autonomy despite Multimorbidity in Saxony through Patient Empowerment: Holistic Care for Older Adults via the Networking of All Regional Facilities and Service Providers.; CHD: coronary heart disease; CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease; mAFA: Mobile Atrial Fibrillation Application; MM: multimorbidity; MASK: Mobile Airways Sentinel Network; MATCH: Multimorbidities Managing Technology for Healthcare; mWellcare: An integrated mHealth System for the Prevention and Care of Chronic Disease; NOMHAD: Novel Model for Health Care Delivery; ProACT: Integrated Technology Systems for Proactive Patient-Centered Care.



The interplay between conditions is operationalized within the projects through different interaction categories. Pathophysiological integration is explicitly addressed in ProACT and MASK, which model biological dependencies between conditions. In contrast, systems like MATCH and ATMoSPHÄRE do not explicitly model these biological synergies. Instead, they manage clinical consequences by reconciling guideline conflicts or integrating multidimensional geriatric assessments to monitor functional decline. MATCH and mWellcare use symptom-based integration by monitoring overlapping clinical markers, such as dyspnea, to detect exacerbations without requiring deep biological modeling.

Treatment-related harmonization is a key feature in ATMoSPHÄRE and mAFA, where conflicting clinical guidelines are reconciled by prioritizing holistic, patient-centered goals. In comparison, while NOMHAD identifies risks, the study does not report any automated tools to adjust treatments. Therefore, based on the published data, clinicians must still resolve therapy conflicts themselves. While ProACT provides holistic self-management support, its treatment-related

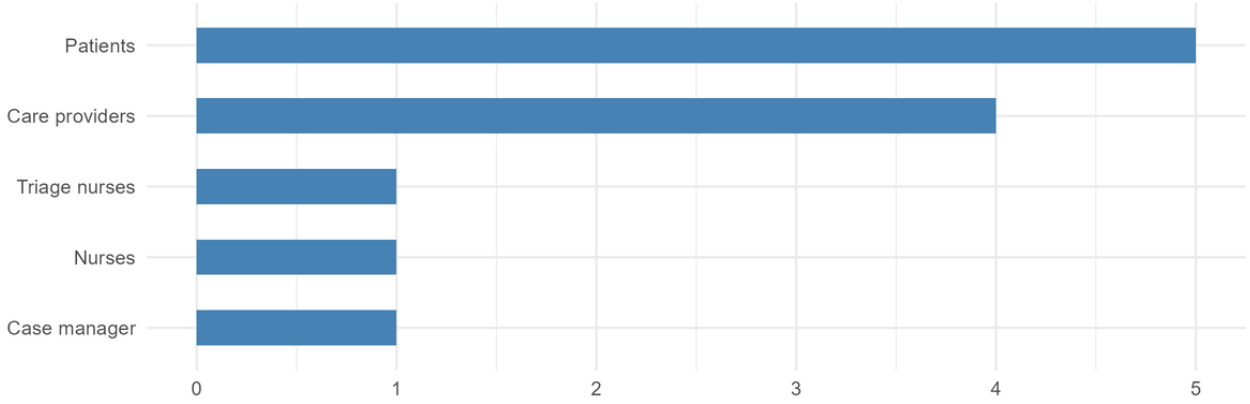
XSL•FO
RenderX

interactions are primarily focused on harmonizing behavioral goals (eg, activity levels) rather than providing automated logic to resolve conflicting clinical guidelines or pharmacological therapies. Finally, all analyzed projects demonstrate strong care process integration, using integrated dashboards or recommendations to support both patient-health care provider communication and interprofessional coordination among the various health care stakeholders involved. The full details and technical justifications of categorization for each project are provided in Multimedia Appendix 5.

Target Groups

The majority of the projects (n=5) are primarily designed for patients, aiming to enhance patient empowerment and promote effective self-management of their health (Figure 7). In addition, four projects specifically target care providers, underscoring the importance of a multiperspective approach in managing multimorbidity. Within this framework, nurses, triage nurses, and case managers frequently emerge as pivotal organizational actors who facilitate coordination and provide ongoing support. Their involvement reflects a model of care that integrates professional guidance with patient autonomy.

Figure 7. Target groups of the included studies (N=11; 2014 - 2024) [25-35].



Evaluation Methods

Of the 11 included studies, nine [25-27,29,31-35] reported an evaluation phase, while the remaining two [28,30] focused on describing the concept and implementation phase. These nine studies varied in terms of design, population, intervention type, and outcome measurement. The most frequently used study design was the proof-of-concept (PoC) trials (n=3), followed by randomized controlled trials (RCTs; n=2) and interventional studies (single-arm design), observational studies, mixed method designs, and pilot studies with one mention each. Both RCTs used cluster randomization approaches to allocate participants at the health center level.

The sample sizes of the studies involving participants (n=9) ranged from eight in ATMoSPHÄRE (2021) to 8504 individuals drawn from The Irish Longitudinal Study on Ageing (TILDA) cohort study used in ProACT (2018), with a median sample size of 120 participants. In total, these studies involved 19,036 individuals, including patients and health care professionals.

Participant demographics across studies consistently reflected an older population except for MASK 2018 (12-92 years; mean 39, SD 16.5 years). Mean participant ages ranged within the other studies from 55.1 years (mWellcare 2019) to 74.2 years (ProACT studies), with several studies reporting SDs of 6 to 13 years. Gender distributions varied: in the mAFA trial, 33.4% of the intervention group were female compared to 41.9% in the control group; in the NOMHAD study, 87% of participants were male. Studies consistently included participants with at least two chronic conditions.

Only two studies [34,35] used control groups. The mAFA-II and mWellcare trials compared outcomes between intervention and control clusters, whereas the remaining studies were single-arm or observational in nature and focused either on feasibility, system development, or user experience.

Outcome measures were diverse. Clinical outcomes included change in systolic blood pressure, hemoglobin A_{1c} (HbA_{1c}) levels, COPD-related quality-of-life scores, and hospitalization rates. Process and behavioral outcomes were also widely reported, such as daily adherence rates to symptom diaries, medication adherence, and digital literacy assessments. User engagement metrics (eg, number of app logins and completion of readings) and usability measures (eg, questionnaire-based satisfaction) were reported in studies such as ProACT 2021 [26] and NOMHAD 2022 [33]. Some studies (eg, ProACT 2018 [25]) focused solely on algorithmic prediction accuracy, reporting performance metrics like Brier scores.

Duration of follow-up or observation ranged from 3 months (eg, NOMHAD 2022 [33]: mean 78.7 days) to 12 months or longer (eg, mAFA-II [34], ProACT [25-27], and MASK 2018 [29]). Most RCTs and PoC trials included clearly defined timelines with multiple data collection points, while observational studies (eg, ATMoSPHÄRE 2021 [32]) were either cross-sectional or did not report exact follow-up durations.

Outcome

The reported outcomes clustered into 4 main areas: user engagement and adherence, usability and implementation,

XSL•FO
RenderX

technical performance and user experience, and clinical effectiveness (Table 2).

Table . Focus and findings of the evaluations reported in the included studies (N=11; 2014 - 2024).

Focus area	Projects	Findings
User engagement and adherence	ProACT[a] [26,27] MATCH[b] [31]	Sustained engagement, moderate adherence, improved patient confidence
Usability and implementation	[c]ATMoSPHÄRE [32], MASK[d] [29]	Usable systems aiding GPs[e], real-world guideline integration, support for allergy care implementation
Technical performance and UX[f]	NOMHAD[g] [33], ProACT [25]	High algorithmic accuracy and patient satisfaction, demonstrated model accuracy using population health data
Clinical outcomes	mAFA-II[h] [34], mWellcare[i] [35]	mAFA-II showed reduced adverse events; mWellcare showed no significant effect

[a]ProACT: Integrated Technology Systems for Proactive Patient-Centered Care.

[b]MATCH: Multimorbidities Managing Technology for Healthcare.

[c]ATMoSPHÄRE: Autonomy despite Multimorbidity in Saxony through Patient Empowerment: Holistic Care for Older Adults via the Networking of All Regional Facilities and Service Providers.

[d]MASK: Mobile Airways Sentinel Network.

[e]GP: general practitioner.

[f]UX: user experience.

[g]NOMHAD: Novel Model for Health Care Delivery.

[h]mAFA: Mobile Atrial Fibrillation Application.

[i]mWellcare: mHealth and Wellness Care.

Studies evaluating patient engagement generally reported sustained use of digital interventions, moderate adherence, and increased confidence in self-management, particularly when technological support was complemented by human assistance. Evaluations of usability and implementation indicated that digital tools could be successfully integrated into routine care and were generally perceived as useful by health care professionals. Studies focusing on technical performance reported high levels of patient satisfaction and promising algorithmic accuracy, highlighting the potential of data-driven approaches for multimorbidity management. Evidence regarding clinical effectiveness was limited and mixed. While the mAFA-II trial demonstrated significant improvements in clinical outcomes, the mWellcare study found no significant intervention effects.

Synthesis of Results

Overview

In the following section, findings are synthesized according to the research objectives to ensure a coherent and structured overview of the evidence.

Technology-Enabled Care for Multimorbidity

The included studies describe integrated, multimodal digital systems for multimorbid care rather than isolated tools. On the patient side, mHealth apps are frequently combined with wearable sensors to enable self-management, continuous symptom tracking, and remote monitoring. For health care providers, data infrastructures and telemedicine platforms support care coordination and information management. These provider solutions often incorporate analytics to transform patient data into actionable insights, such as risk indicators and multimorbidity patterns. While some systems only provide basic dashboards and alerts, others use decision support tools and decision recommendation systems to offer personalized, evidence-based recommendations for complex patient profiles.

Addressing Disease Interactions in Intervention Design

The included studies show differences in how multimorbidity-aware functionality is implemented across systems. Some approaches focus on analytics that identify multimorbidity-related patterns, generate risk profiles, and, in certain cases, explore interactions between conditions. Other systems emphasize decision support through context-sensitive dashboards and alerting mechanisms that highlight clinically relevant risks and support decision-making in complex patient situations. In more advanced cases, interventions provide prescriptive decision support by integrating clinical guidelines, resolving potential conflicts, and generating individualized recommendations.

Regarding disease interactions, most interventions focused on symptom-level and patient-reported interactions, using composite scores, symptom diaries, or patient-reported outcome measures to capture cross-disease burden. Care process interactions were also frequently addressed through dashboards, risk indicators, and coordinated care workflows supporting prioritization and decision-making across conditions. In contrast, only a few interventions incorporated more advanced pathophysiological modeling, such as BN-based analytics in ProACT, or systematically addressed treatment-related interactions through guideline-based decision support. Overall, current digital health interventions tend to aggregate and relate multimorbidity data, but they rarely operationalize deeper

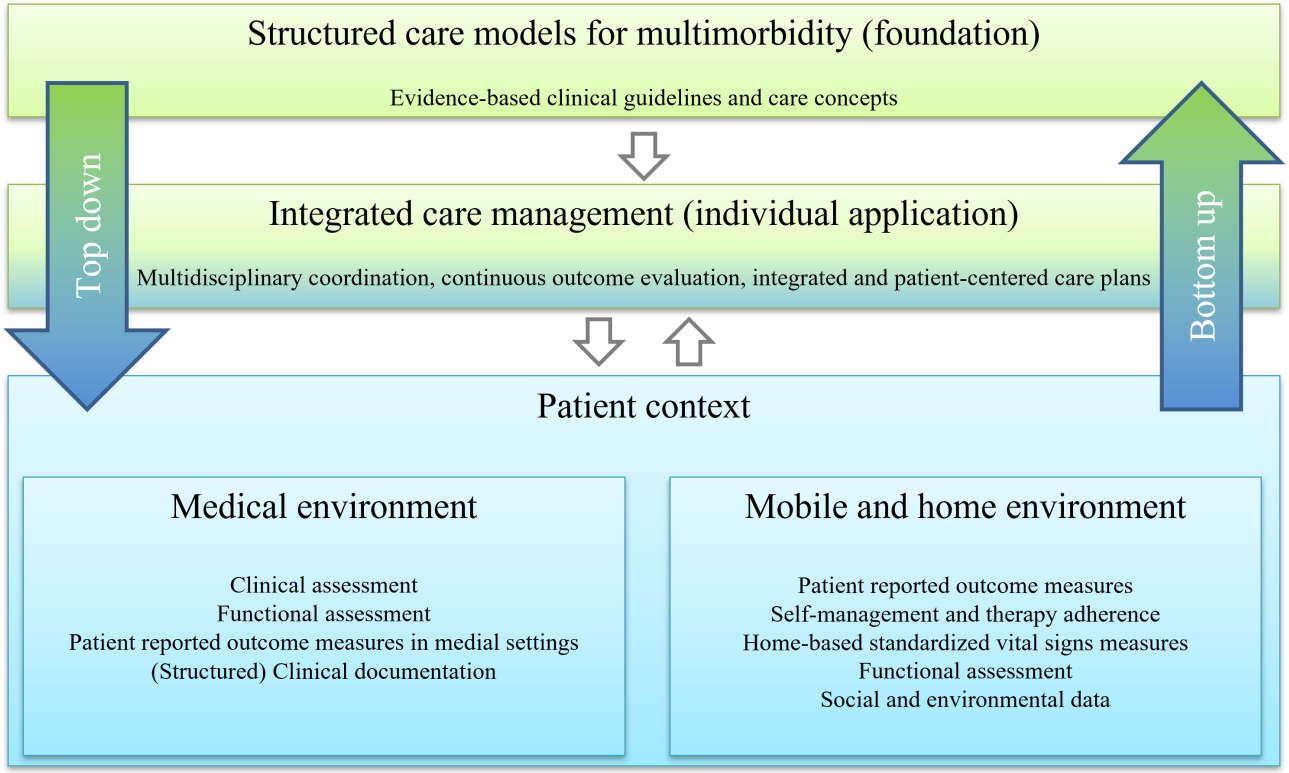
XSL•FO
RenderX

disease interdependencies or explicitly reconcile conflicting treatment recommendations.

Concepts for Integrated, Personalized Multimorbidity Care

The involved studies reflect a multidimensional approach to managing multimorbidity, integrating top-down, guideline-driven strategies and bottom-up, patient-centered data-driven models. The findings can be organized into three complementary domains (Figure 8): (1) structured care models for multimorbidity, (2) integrated care management, and (3) standardized documentation of patient health profiles.

Figure 8. Conceptual overview of multimorbidity care approaches: Synthesis of components for managing disease interactions across included projects (n=7) based on a review of 11 included studies (2014 - 2024). [25-35].



The first domain is based on clinical guidelines covering both disease-specific recommendations and multimorbidity-oriented frameworks. The ABC pathway and the Allergic Rhinitis and its impact on Asthma (ARIA) 2015 guideline demonstrate this by promoting holistic, integrated care across coexisting conditions. Embedding such guidelines into digital interventions allows for automated, evidence-based recommendations for complex patient profiles. Integrated care management relies on structured programs that coordinate roles, tools, and care plans across providers. Defined roles (triage nurses and case managers) guide patients through complex treatment pathways, while digital interventions help translate evidence-based protocols into clinical practice.

Standardized documentation uses validated scoring and assessment instruments to describe complex patient profiles, identify individual priorities, and support dynamic adjustment of care. These include disease-specific risk scores such as Congestive heart failure, Hypertension, Age, Diabetes mellitus, Stroke/transient ischemic attack/thromboembolism, Vascular disease, Age, Sex category (CHA_2DS_2); Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly, Drugs/alcohol concomitantly (HAS-BLED); and Sex, Age, Medical history, Treatment (interacting drugs), Tobacco use, Race (non-Caucasian; SAMe-TT_2R_2); SAMe-TT_2R_2 in mAFA-II; the ORI composite index in NOMHAD; and geriatric assessments such as the Instrumental Activities of Daily Living (IADL) scale, Mini-Mental State Examination, and Timed Up-and-Go in ATMoSPHÄRE. Patient-reported outcomes including symptom diaries, quality-of-life questionnaires, and standardized tools such as the International Physical Activity Questionnaire (IPAQ) and Control of Allergic Rhinitis and Asthma Test (CARAT) support active self-monitoring alongside clinician-administered assessments. Continuous vital sign tracking via wearables and visual analog scales (VAS) in MASK further extend data collection. Together, these instruments support automated risk stratification and feed analytical modules to enable adaptive, need-based treatment planning.

Frequently Targeted Diseases and Recurring Multimorbidity Clusters

The analyzed studies targeted various multimorbidity clusters, including cardiometabolic multimorbidity, cardiorespiratory multimorbidity, and allergy-asthma cluster. These clusters predominantly consist of chronic conditions, rather than acute illness, thus addressing long-term treatment and management. The multimorbidity clusters are characterized by overlapping symptoms and complex disease interactions between conditions, which complicate treatment. These challenges are addressed by

XSL•FO
RenderX

digital solutions examined in this review. For example, we identified three projects (ProAct, MATCH, and NOMHAD) that focus on the cooccurrence of COPD and CHF, partly in combination with other diseases, highlighting the importance of managing overlapping symptoms, shared risk factors, therapeutic overlaps, and pathophysiological interactions of these two diseases. There is only one project that targets a multimorbidity cluster (allergy-asthma) also relevant to the younger population. Apart from this project, we found others focusing also on younger populations, which we excluded because the digital solutions address only a single disease in patients with multimorbidities. However, there is clear evidence that multimorbidity also affects younger populations (eg, addiction-depression cluster), but multimorbidity-focused approaches for the younger population remain insufficiently addressed in current digital health solutions.

Reported Effectiveness and Implementation Outcomes of Multidisease Approaches

The reviewed studies demonstrate that digital health interventions targeting patients with multiple chronic conditions can be both feasible to implement and potentially effective. Successful implementation was generally associated with high usability, adaptability to clinical settings, and the integration of personal support mechanisms.

Evidence suggests that effectiveness in multidisease contexts depends not only on the technological capabilities of the intervention but also on how well it is embedded within broader care pathways. Interventions that combined digital tools with human support, such as health professionals assisting with data interpretation, tended to show stronger engagement and sustained use over time.

While one reported measurable clinical benefit, another highlighted implementation success without clear clinical superiority over usual care. This variability points to the importance of context, user engagement, and implementation fidelity in shaping outcomes.

Overall, the findings support the potential of multidisease digital approaches, especially when they are designed to be user-centered, support hybrid care models, and align with existing health care infrastructures. However, further research is needed to understand the conditions under which these interventions deliver consistent clinical value at scale.

Challenges and Barriers to the Real-World Application of Interventions

While several studies mention challenges, these are often described only in general terms or focus on different aspects, without directly addressing our specific research question. Overall, the field appears to be in its early stages, with most of the existing literature limited to feasibility or explorative research. As such, critical steps toward the broader implementation and scaling of these interventions remain largely unaddressed. Consequently, our guiding question regarding the reported challenges, limitations, or barriers affecting the practical application, scalability, or sustainability of these interventions in real-world settings cannot yet be adequately answered, as current research has not progressed to this stage.

Discussion

Summary of Evidence

This scoping review identified only a small number of digital health interventions that explicitly address multimorbidity as an interconnected phenomenon rather than a collection of separate chronic diseases. Although several interventions supported patients with multiple chronic conditions, only a few implemented mechanisms to model disease interactions or reconcile competing treatment recommendations.

This reflects broader challenges in multimorbidity care, where health care structures and clinical guidelines continue to be organized around individual diseases [5,6,8,11,13]. Several factors may explain why multimorbidity-aware digital interventions remain limited in practice. First, health care systems and clinical guidelines are still predominantly organized around single diseases, making integrated digital modeling of multiple conditions inherently difficult [5,6,8,13]. Second, multimorbidity introduces substantial technical and clinical complexity, including overlapping symptoms, competing treatment recommendations, heterogeneous patient trajectories, and high data integration requirements [8,11,37]. Third, the development and evaluation of such systems often require interdisciplinary collaboration across clinical, technical, and organizational domains, which may limit scalability and implementation in routine care settings [12,13]. Together, these challenges may contribute to the predominance of symptom-level monitoring and care coordination approaches over more advanced multimorbidity modeling strategies.

Although theoretical models for integrated multimorbidity care already exist, their digital translation remains limited. Established clinical and public health frameworks, including patient-centered care models, provide conceptual direction, yet current eHealth systems rarely operationalize them in practice [8,12,38]. Many interventions aggregate information across conditions without explicitly modeling their interdependencies or supporting integrated therapeutic decision-making. Similar fragmentation has also been described in traditional care settings, where established links between diseases, such as diabetes and periodontitis, are often recognized clinically but not jointly managed [39].

Beyond summarizing the available evidence, this review provides a structured perspective on multimorbidity-aware digital systems by distinguishing both the functional depth of technological support (analytics, dashboards, and recommendation systems) and the types of disease interactions addressed (pathophysiological, treatment-related, symptom-related, and care-process interactions). This dual perspective may support future development and evaluation of multimorbidity-oriented digital interventions by providing a common framework for describing system capabilities.

The included studies highlight important facilitators for successful implementation. Interventions combining remote monitoring, self-management support, data analytics, and structured care roles (eg, triage nurses and case managers) demonstrated that successful multimorbidity care depends not

XSL•FO
RenderX

only on technical functionalities but also on organizational integration and human support [25-33]. However, even these interventions primarily addressed symptom-level and care-process integration rather than deeper pathophysiological or treatment-related interactions.

Overall, digital health for multimorbidity is still in an early developmental phase, in which conceptual models, technical capabilities, and care structures are not yet fully aligned. Most evidence therefore remains limited to feasibility, usability, and implementation outcomes rather than clinical effectiveness. Future research should focus on interoperable multimorbidity-aware infrastructures that integrate heterogeneous patient data, support coordinated care pathways, and investigate how disease interactions can be represented computationally and translated into transparent, clinically actionable decision support. In addition, clinical effectiveness needs to be systematically evaluated to build stronger evidence.

Limitations

This review has several limitations that should be acknowledged. First, the literature search was limited to two databases (PubMed and CINAHL) and to publications in English and German. Although this may have resulted in relevant studies being missed, backward reference screening of included studies and relevant reviews identified only one additional eligible study, suggesting that the risk of substantial omission is limited.

Second, this review deliberately applied strict eligibility criteria by including only interventions that explicitly addressed interactions between coexisting conditions. Consequently, digital health interventions targeting multimorbid populations but managing conditions independently, as well as user-centered design studies, technical development papers, and conceptual frameworks without implementation or evaluation, were excluded. The findings should therefore be interpreted as representing a specific subset of multimorbidity-oriented digital interventions rather than the broader landscape of digital health for multimorbid populations.

Third, only 11 studies [25-35] met the strict inclusion criteria. While this reflects the emerging state of research in multimorbidity-aware digital health, it also limits the generalizability of the findings. In addition, the considerable heterogeneity of study designs, intervention types, and reported outcomes precluded formal comparisons of effectiveness.

Finally, the analytical framework developed in this review to classify multimorbidity-aware system functionalities and disease interaction types was derived inductively from the included studies and informed by existing conceptual literature. Although it provides a structured perspective for describing and comparing multimorbidity-oriented digital interventions, it has not yet been externally validated and should therefore be regarded as a conceptual framework rather than an established taxonomy.

Conclusions

Advancing digital health for multimorbidity will require a transition from disease-oriented technologies toward interoperable, patient-centered systems capable of integrating heterogeneous clinical, behavioral, and contextual data across conditions. Beyond technological innovation, successful implementation will depend on the alignment of digital infrastructures with integrated care models, multidisciplinary workflows, and supportive health care policies.

As multimorbidity increasingly becomes the dominant reality of chronic care, the development of multimorbidity-aware digital systems may become a key prerequisite for scalable and sustainable health care delivery.

Acknowledgments

The authors declare the use of generative AI in the research and manuscript preparation process. A general-purpose large language model (ChatGPT, OpenAI) was used to support terminology expansion during the development of the search strategy and to assist with language refinement and formulation of parts of the manuscript text. All generated suggestions and formulations were critically reviewed, revised, and approved by the authors. The LLM was not used for study screening, data extraction, data analysis, or independent interpretation of findings. The authors take full responsibility for the content of this manuscript.

Funding

The authors received funding support from the Bavarian Center for Digital Health and Social Care for this publication.

Authors' Contributions

Conceptualization: KB, SH
Formal analysis: KB, SH
Investigation: KB, SH
Methodology: KB, SH
Project administration: KB
Writing – original draft: All authors
Writing – review & editing: All authors
All authors approved the final version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final search strings for each database.
[DOC File, 823 KB - ojphi_v18i1e88021_app1.doc]

Multimedia Appendix 2
Detailed overview of data items for data extraction.
[DOC File, 105 KB - ojphi_v18i1e88021_app2.doc]

Multimedia Appendix 3
Study-specific data extraction (study design, sample size, and outcomes).
[DOC File, 133 KB - ojphi_v18i1e88021_app3.doc]

Multimedia Appendix 4
Data extraction on project level (addressed diseases, target group, technologies, functionalities, and data categories).
[DOC File, 115 KB - ojphi_v18i1e88021_app4.doc]

Multimedia Appendix 5
Reported functional depth and interaction categories of the included projects (n=7, reported in 11 studies, 2014-2024).
[DOC File, 134 KB - ojphi_v18i1e88021_app5.doc]

Checklist 1
PRISMA-ScR Checklist
[DOC File, 153 KB - ojphi_v18i1e88021_app6.doc]

References

1. Moffat K, Mercer SW. Challenges of managing people with multimorbidity in today's healthcare systems. BMC Fam Pract 2015 Oct 14;16(16):129. [doi: 10.1186/s12875-015-0344-4] [Medline: 26462820]
2. Cummings M, Bradley J, Teal G. Patient co-design of digital health storytelling tools for multimorbidity: a phenomenological study. Health Expect 2022 Dec;25(6):3073-3084. [doi: 10.1111/hex.13614] [Medline: 36165104]
3. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. EClinicalMedicine 2023 Mar;57:101860. [doi: 10.1016/j.eclinm.2023.101860] [Medline: 36864977]
4. "Just one thing after another" living with multiple conditions. : Taskforce on Multiple Conditions; 2018 Oct 10 URL: https://www.richmondgroupofcharities.org.uk/publications/just-one-thing-after-another-living-with-multiple-conditions/ [accessed 2026-08-13]
5. Hopkins R, Shaver K, Weinstock RS. Management of adults with diabetes and cognitive problems. Diabetes Spectr 2016 Nov;29(4):224-237. [doi: 10.2337/ds16-0035] [Medline: 27899874]
6. Teck J. Diabetes-associated comorbidities. Prim Care 2022 Jun;49(2):275-286. [doi: 10.1016/j.pop.2021.11.004]
7. Pretorius S, Bartle V, Bellass S, et al. Making clinical guidelines work for people with multiple long term conditions: analysis and recommendations from review of single condition guidelines. BMJ Med 2026;5(1):e001495. [doi: 10.1136/bmjmed-2025-001495] [Medline: 41756103]
8. Abdi S, Spann A, Borilovic J, de Witte L, Hawley M. Understanding the care and support needs of older people: a scoping review and categorisation using the WHO international classification of functioning, disability and health framework (ICF). BMC Geriatr 2019 Jul 22;19(1):195. [doi: 10.1186/s12877-019-1189-9] [Medline: 31331279]
9. EHealth: report by the secretariat. : World Health Organization; 2005 Apr 7 URL: https://iris.who.int/items/3fe3e0e6-60f2-4903-a3aa-468ff97d4f0f [accessed 2026-08-13]
10. Lang C, Voigt K, Neumann R, Bergmann A, Holthoff-Detto V. Adherence and acceptance of a home-based telemonitoring application used by multi-morbid patients aged 65 years and older. J Telemed Telecare 2022 Jan;28(1):37-51. [doi: 10.1177/1357633X20901400] [Medline: 32009577]
11. Freilich J, Nilsson GH, Ekstedt M, Flink M. "Standing on common ground" - a qualitative study of self-management support for patients with multimorbidity in primary health care. BMC Fam Pract 2020 Nov 17;21(1):233. [doi: 10.1186/s12875-020-01290-y] [Medline: 33203401]

XSL•FO
RenderX

12. Macdonald GG, Townsend AF, Adam P, et al. eHealth technologies, multimorbidity, and the office visit: qualitative interview study on the perspectives of physicians and nurses. J Med Internet Res 2018 Jan 26;20(1):e31. [doi: 10.2196/jmir.8983] [Medline: 29374004]
13. Rijken M, Struckmann V, Heide I, et al. In: Richardson E, Ginneken E, editors. How to Improve Care for People with Multimorbidity in Europe?: European Observatory on Health Systems and Policies; 2017. [Medline: 29144712]
14. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. J Am Med Inform Assoc 2020 Jun 1;27(6):939-945. [doi: 10.1093/jamia/ocaa029] [Medline: 32357368]
15. Holmes NA, van Agteren JE, Dorstyn DS. A systematic review of technology-assisted interventions for co-morbid depression and substance use. J Telemed Telecare 2019 Apr;25(3):131-141. [doi: 10.1177/1357633X17748991] [Medline: 29310532]
16. Schouten MJE, Christ C, Dekker JJM, Riper H, Goudriaan AE, Blankers M. Digital interventions for people with co-occurring depression and problematic alcohol use: a systematic review and meta-analysis. Alcohol Alcohol 2022 Jan 8;57(1):113-124. [doi: 10.1093/alcalc/agaa147] [Medline: 33588432]
17. Kallas D, Sandhu N, Gandilo C, et al. Use of digital health technology in heart failure and diabetes: a scoping review. J Cardiovasc Transl Res 2023 Jun;16(3):526-540. [doi: 10.1007/s12265-022-10273-6] [Medline: 35639339]
18. Shah A, Hussain-Shamsy N, Strudwick G, Sockalingam S, Nolan RP, Seto E. Digital health interventions for depression and anxiety among people with chronic conditions: scoping review. J Med Internet Res 2022 Sep 26;24(9):e38030. [doi: 10.2196/38030] [Medline: 36155409]
19. Cajamarca G, Proust V, Herskovic V, Cádiz RF, Verdezoto N, Fernández FJ. Technologies for managing the health of older adults with multiple chronic conditions: a systematic literature review. Healthcare (Basel) 2023 Nov 3;11(21):2897. [doi: 10.3390/healthcare11212897] [Medline: 37958041]
20. Maehder K, Löwe B, Härter M, Heddaeus D, Scherer M, Weigel A. Management of comorbid mental and somatic disorders in stepped care approaches in primary care: a systematic review. Fam Pract 2019 Jan 25;36(1):38-52. [doi: 10.1093/fampra/cmy122] [Medline: 30535053]
21. Wiwatkunupakarn N, Aramrat C, Pliannuom S, et al. The integration of clinical decision support systems into telemedicine for patients with multimorbidity in primary care settings: scoping review. J Med Internet Res 2023 Jun 28;25:e45944. [doi: 10.2196/45944] [Medline: 37379066]
22. Kraef C, van der Meirschen M, Free C. Digital telemedicine interventions for patients with multimorbidity: a systematic review and meta-analysis. BMJ Open 2020 Oct 13;10(10):e036904. [doi: 10.1136/bmjopen-2020-036904] [Medline: 33051232]
23. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
24. Pollock D, Peters MDJ, Khalil H, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 1;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
25. Deparis S, Pascale A, Tommasi P, Kotoulas S. An analytical method for multimorbidity management using Bayesian networks. Stud Health Technol Inform 2018;247(247):820-824. [doi: 10.3233/978-1-61499-852-5-820] [Medline: 29678075]
26. Doyle J, Murphy E, Gavin S, et al. A digital platform to support self-management of multiple chronic conditions (ProACT): findings in relation to engagement during a one-year proof-of-concept trial. J Med Internet Res 2021 Dec 15;23(12):e22672. [doi: 10.2196/22672] [Medline: 34914612]
27. Doyle J, McAleer P, van Leeuwen C, et al. The role of phone-based triage nurses in supporting older adults with multimorbidity to digitally self-manage - findings from the ProACT proof-of-concept study. Digit Health 2022;8(8):20552076221131140. [doi: 10.1177/20552076221131140] [Medline: 36238753]
28. Bousquet J, Schunemann HJ, Fonseca J, et al. MACVIA-ARIA Sentinel NetworK for allergic rhinitis (MASK-rhinitis): the new generation guideline implementation. Allergy 2015 Nov;70(11):1372-1392. [doi: 10.1111/all.12686] [Medline: 26148220]
29. Bousquet J, Devillier P, Anto JM, et al. Daily allergic multimorbidity in rhinitis using mobile technology: A novel concept of the MASK study. Allergy 2018 Aug;73(8):1622-1631. [doi: 10.1111/all.13448] [Medline: 29569295]
30. Bousquet J, Arnavielhe S, Bedbrook A, et al. MASK 2017: ARIA digitally-enabled, integrated, person-centred care for rhinitis and asthma multimorbidity using real-world-evidence. Clin Transl Allergy 2018;8(1):45. [doi: 10.1186/s13601-018-0227-6] [Medline: 30386555]
31. Sloots J, Bakker M, van der Palen J, et al. Adherence to an eHealth self-management intervention for patients with both COPD and heart failure: results of a pilot study. Int J Chron Obstruct Pulmon Dis 2021;16(16):2089-2103. [doi: 10.2147/COPD.S299598] [Medline: 34290502]
32. Borchers P, Piller S, Böhme M, Voigt K, Bergmann A. Need-based care of multi-morbid patients - supporting general practitioners with algorithm-generated recommendations of healthcare services (telemedicine-project ATMoSPHÄRE). BMC Fam Pract 2021 Oct 8;22(1):198. [doi: 10.1186/s12875-021-01537-2] [Medline: 34625053]
33. Chacornac M, Faoro A, Texereau J, Billoet C, Hominal S. Performance of an eHealth (NOMHAD) system comprising telemonitoring, telenotification, and telecoaching for patients with multimorbidity: proof-of-concept study. JMIR Form Res 2022 Mar 11;6(3):e32205. [doi: 10.2196/32205] [Medline: 35275068]

XSL•FO
RenderX

34. Yao Y, Guo Y, Lip GYH, mAF-App II Trial investigators. The effects of implementing a mobile health–technology supported pathway on atrial fibrillation–related adverse events among patients with multimorbidity. JAMA Netw Open 2021 Dec 1;4(12):e2140071. [doi: 10.1001/jamanetworkopen.2021.40071] [Medline: 34932104]
35. Prabhakaran D, Jha D, Prieto-Merino D, et al. Effectiveness of an mHealth-based electronic decision support system for integrated management of chronic conditions in primary care: the mWellcare cluster-randomized controlled trial. Circulation 2019 Jan 15;139(3):380-391. [doi: 10.1161/CIRCULATIONAHA.118.038192] [Medline: 30586732]
36. Mehl G, Tamrat T, Labrique A, et al. Classification of digital health interventions v 1.0. : World Health Organization; 2018 Apr. [doi: 10.13140/RG.2.2.14531.30243]
37. O'Sullivan D, Woensel WV, Wilk S, et al. Towards a framework for comparing functionalities of multimorbidity clinical decision support: a literature-based feature set and benchmark cases. AMIA Annu Symp Proc 2021;2021:920-929. [Medline: 35308994]
38. Mangin D, Parascandalo J, Khudoyarova O, Agarwal G, Bismah V, Orr S. Multimorbidity, eHealth and implications for equity: a cross-sectional survey of patient perspectives on eHealth. BMJ Open 2019 Feb 12;9(2):e023731. [doi: 10.1136/bmjopen-2018-023731] [Medline: 30760515]
39. Seitz MW, Haux C, Smits KPJ, et al. Development and evaluation of a mobile patient application to enhance medical-dental integration for the treatment of periodontitis and diabetes. Int J Med Inform 2021 Aug;152(152):104495. [doi: 10.1016/j.ijmedinf.2021.104495] [Medline: 34010786]

Abbreviations

ABC: A: anticoagulation or avoid stroke; B: better symptom control; and C: cardiovascular disease and comorbidity management
AF: atrial fibrillation
ARIA : allergic rhinitis and its impact on asthma
ATMoSPHÄRE: Autonomy despite Multimorbidity in Saxony through Patient Empowerment: Holistic Care for Older Adults via the Networking of All Regional Facilities and Service Providers
BN: Bayesian Network
CARAT: Control of Allergic Rhinitis and Asthma Test
CDSS: clinical decision support system
CHA2DS2: Congestive heart failure, Hypertension, Age, Diabetes mellitus, Stroke/transient ischemic attack/thromboembolism, Vascular disease, Age, Sex category
CHD: coronary heart disease
CHF: congestive heart failure
COPD: chronic obstructive pulmonary disease
DRS: digital recommendation system
DSS: decision support system
HAS-BLED: Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly, Drugs/alcohol concomitantly
HbA_{1c}: hemoglobin A_{1c}
IADL: Instrumental Activities of Daily Living
IPAQ: International Physical Activity Questionnaire
LLM: large language model
mAFA: Mobile Atrial Fibrillation Application
MASK: Mobile Airways Sentinel Network
MATCH: Multimorbidities Managing Technology for Healthcare
mHealth: mobile health
mWellcare: mHealth and Wellness Care
NOMHAD: Novel Model for Health Care Delivery
ORI: overall risk indicator
PCC: Population, Concept, and Context
PoC: proof of concept
PRISMA-S : Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
PRISMA-ScR : Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
ProACT: Integrated Technology Systems for Proactive Patient-Centered Care
RCT: randomized controlled trial
TILDA: The Irish Longitudinal Study on Ageing
VAS: visual analog scale
WHO: World Health Organization

XSL•FO
RenderX

Edited by E Mensah; submitted 18.Nov.2025; peer-reviewed by J Roca, Y An; revised version received 30.Jun.2026; accepted 27.Jul.2026; published 08.Sep.2026.

Please cite as:
Bochtler K, Houta S
eHealth Approaches to Multimorbidity Management: Scoping Review of Interventions and Multidisease Strategies
Online J Public Health Inform 2026;18:e88021
URL: https://ojphi.jmir.org/2026/1/e88021
doi:10.2196/88021

© Katja Bochtler, Salima Houta. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 8.Sep.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Associations Between Hospital Structural Characteristics and Adoption of Public Health Data Integration and Automation: National Cross-Sectional proofsStudy

Hanadi Hamadi, PhD; Alaysia Alford, MHA; Chloe Smith Lopez, MHA; Durron Baker, MHA; James Ian Samaniego, MHA; Jessica Yu Jin, MHA

Department of Health Administration, University of North Florida, 1 UNF Drive, Jacksonville, FL, United States

Corresponding Author:
Hanadi Hamadi, PhD
Department of Health Administration, University of North Florida, 1 UNF Drive, Jacksonville, FL, United States

Abstract

Background: Public health data integration and automation systems are crucial for effective health care delivery and public health surveillance. However, the factors associated with hospitals' adoption and successful implementation remain inadequately explored.

Objective: This study aims to examine how hospital characteristics influence the adoption of public health data integration and automation.

Methods: We analyzed 2277 hospitals from the 2023 American Hospital Association Annual Survey and its Health Information Technology supplement, focusing on 6 public health reporting categories. Multivariable logistic regression models were used to examine the association between hospital characteristics and the 2 primary outcomes: active electronic data submission and use of automated transmission processes.

Results: System-affiliated and not-for-profit hospitals demonstrated significantly higher rates of electronic data submission and automated reporting across most categories (odds ratio [OR] 1.70 - 2.27; $P<.001$). Rural hospitals showed lower adoption rates in immunization registry (OR 0.77, 95% CI 0.61-0.97), public health registry (OR 0.67, 95% CI 0.46-0.97), and clinical data registry reporting (OR 0.77, 95% CI 0.60-0.98). Larger hospitals were more likely to implement electronic reporting, with medium and large hospitals showing stronger engagement in syndromic surveillance reporting (OR 1.52, 95% CI 1.06-2.19 and OR 2.29, 95% CI 1.17-4.46, respectively). Teaching status was significantly associated only with clinical data registry reporting (OR 2.66, 95% CI 1.56-4.52 for major teaching hospitals).

Conclusions: Hospital characteristics, particularly system affiliation, ownership type, and geographic location, are strongly associated with public health data integration and automation capabilities. Findings suggest targeted interventions are needed to address disparities in smaller and rural facilities to ensure equitable advancement of public health reporting infrastructure.

(Online J Public Health Inform 2026;18:e86263) doi:10.2196/86263

KEYWORDS

public health data; data reporting; automation; active reporting; hospitals

Introduction

The integration and automation of public health data have evolved from manual record-keeping to modern digital systems that enhance real-time data sharing and interoperability. Automated frameworks now combine structured and unstructured health data, improving research capabilities and public health responsiveness. The implementation of Findable, Accessible, Interoperable, and Reusable data principles has further enhanced data use for decision-making [1]. These innovations highlight the importance of technology-driven data integration in optimizing health care delivery and public health outcomes [2]. The Health Information Technology for Economic and Clinical Health Act of 2009, enacted into law by Title XIII of the American Recovery and Reinvestment Act of 2009, dramatically fueled computerization in health care through reimbursement incentives to adopt electronic health records (EHRs) as a method of standardizing and enhancing interoperability of data [3,4]. These differences reiterated that institutional resilience and organizational readiness were more critical than technology availability to successful adoption.

The 2020s have seen further advancements with artificial intelligence (AI) and machine learning technologies extending automation capabilities. AI-based software now streamlines tasks such as drug safety compliance reporting, reducing administrative burdens and human error [5]. Despite technological progress, persistent challenges remain:

XSL•FO
RenderX

interoperability gaps prevent smooth data exchange between institutions due to diverse standards and proprietary tools [6]; regulatory requirements often fail to address structural barriers such as system upgrade costs or personnel training needs [5]; and workforce preparedness is frequently overlooked, particularly in low-resource settings where staff may lack proper training to use new technologies effectively [7].

Public reporting of hospital data, such as patient outcomes, infection rates, and readmission rates, can drive improvements in health care quality by promoting transparency and accountability. Studies have shown that hospitals participating in public reporting programs tend to engage in quality improvement activities more actively [8,9]. For instance, the American College of Cardiology's voluntary public reporting program revealed that hospitals with higher participation rates demonstrated better performance in cardiac care [8,9]. Additionally, the COVID-19 pandemic highlighted the importance of standardized and automated reporting systems to ensure timely and accurate data exchange, which is essential for effective public health responses and leads to better health outcomes for patients [4].

The association between public health data integration, automation, and hospital characteristics has become a key focus in assessing reporting system effectiveness, particularly during the COVID-19 pandemic [10]. Beyond improving health care delivery, data integration can enhance hospitals' operational efficiency, potentially leading to higher profits and increased patient service capacity [11]. However, the most significant barrier to integration remains the lack of standardization in health data norms at local, national, and international levels. Many health data systems cannot communicate effectively, resulting in integration challenges when patients move between health systems [12]. Through improved data integration, public health systems can better address concerns like social determinants of health and disease monitoring for future pandemics while enhancing patient experiences through personalized care. While prior research has examined EHR adoption broadly, few studies have disaggregated public health reporting into its component categories to identify differential adoption patterns across hospital characteristics. This study addresses this gap by simultaneously examining 6 distinct public health reporting categories and analyzing both electronic submission engagement and automation processes as separate outcomes. This context situates the central question of this research: *What hospital characteristics are associated with the adoption and success of automated health reporting systems?* By identifying factors associated with successful implementation of automated health reporting systems, the findings can inform strategies to address disparities and improve public health data infrastructure across different health care settings. This research is particularly significant in light of the COVID-19 pandemic, which exposed weaknesses in current health data systems, especially regarding integration and automation [10]. A well-integrated, automated health data system will not only lead to improved patient outcomes and more patient-focused care but also enhance public health decision-making at both local and national levels [13].

Methods

Data Source

The primary data for this study were derived from the 2023 American Hospital Association (AHA) Annual Survey and its supplemental Health Information Technology Survey [14]. The AHA Annual Survey provides comprehensive information on a wide range of hospital characteristics including organizational structure, service lines, staffing, finances, and patient populations. The supplemental Health Information Technology Survey specifically captures detailed information about hospitals' health information technology capabilities, EHR implementation, and public health reporting practices.

Outcome Variables

The first set of outcome variables assessed the hospital's current stage of active engagement towards electronically submitting data for public health reporting across 6 categories: syndromic surveillance, immunization registry, electronic case reporting, public health registry, clinical data registry, and electronic reportable laboratory result reporting. For each category, respondents selected one of five ordinal response options representing implementation stages: (1) actively electronically submitting production data, (2) in the process of testing and validating electronic submission, (3) completed registration to submit data, (4) have not completed registration, or (5) do not know. This variable was operationalized as a dichotomous (yes or no) measure, with "yes" representing hospitals that reported actively electronically submitting production data and those that did not (yes or no). This dichotomization approach was used to create a clear distinction between hospitals actively engaged in electronic reporting versus those at earlier implementation stages or nonparticipants, consistent with prior AHA survey analyses examining health IT adoption [15].

The second set of outcome variables assessed the specific processes used to transmit health data, with respondents identifying whether their hospital utilized automated, manual, or mixed processes across 6 reporting categories. Response options included: (1) fully or primarily automated, (2) mix of automated and manual processes, (3) fully or primarily manual, or (4) do not know. For analysis purposes, the automated reporting variable was operationalized as a binary (yes or no) measure for each of the seven reporting categories, with "yes" representing hospitals using fully or primarily automated processes.

Confounding Variables

The analysis also included several hospital characteristics and market factors that may influence public health data reporting practices. Hospital ownership type was categorized as government (federal and nonfederal), not-for-profit (private hospitals with Internal Revenue Service 501(c)(3) tax-exempt status), or for-profit (investor-owned facilities operating as taxable business entities). Geographic location was classified as rural or nonrural (urban) based on the hospital's physical setting and Rural-Urban Commuting Area codes. Hospital size was operationalized using the total staffed bed count and

XSL•FO
RenderX

stratified into 3 categories: small (fewer than 100 beds), medium (100 - 299 beds), and large (300 or more beds).

System affiliation was measured as a binary variable indicating whether the hospital was part of a larger health care system (system-affiliated) or operated independently. Teaching status was classified using the AHA criteria into nonteaching or teaching. Medicare percentage (proportion of total Medicare inpatient days) and Medicaid percentage (proportion of total Medicaid inpatient visits) were included to account for patient population characteristics that may influence hospitals' priorities and resource allocation for health IT investments. Market competition was measured using the Herfindahl-Hirschman Index (HHI), calculated based on the distribution of hospital beds within each health care market area. Higher HHI values indicate greater market concentration and less competition, with values approaching 1.0 representing highly concentrated markets [16]. This measure was included to control for the potential influence of competitive pressures on hospitals' public health reporting practices and technology adoption decisions. These variables were selected based on previous literature identifying them as potential determinants of health care technology adoption, organizational innovation, and public health reporting capabilities.

Statistical Analysis

This study used descriptive statistics and logistic regression analyses. For categorical variables, we computed frequencies and percentages. For continuous variables (Medicare percentage, Medicaid percentage, and HHI), we calculated means and SDs. We stratified these descriptive statistics by our two primary outcome measures: (1) whether hospitals were actively submitting data electronically and (2) whether hospitals used automated processes for data transmission.

For our primary analysis, we developed a series of multivariable logistic regression models to examine the adjusted associations between hospital characteristics and public health reporting practices. Separate models were constructed for each of the 6 reporting categories (syndromic surveillance, immunization registry, electronic case reporting, public health registry, clinical data registry, and electronic reportable laboratory result reporting) and for both outcome measures (active electronic submission and automated processes).

Results from the logistic regression models are presented as adjusted odds ratios (ORs). We conducted model diagnostics to ensure that all logistic regression assumptions were met. These included tests for multicollinearity using variance inflation factors, examination of influential observations using the Cook distance, and assessment of model fit using the Hosmer-Lemeshow goodness-of-fit test. All analyses were conducted using Stata (version 17.0; StataCorp), with statistical significance set at $P<.05$ for all tests. Cases with missing data on any study variables were excluded from the analysis using listwise deletion.

Ethical Considerations

In accordance with the policy of the university of North Florida, the Institutional Review Board for the Protection of Human Subjects categorized the research as exempt since the study analyzed secondary data that are publicly available.

Results

The results reveal patterns in electronic health data reporting practices between health care facilities based on their patient demographics, market concentration measures, and hospital characteristics.

Actively Submitting Data Electronically

Table 1 reports hospital categorical characteristics across hospitals that actively submit data electronically versus those that do not.

Table 2 reports hospital and market continuous characteristics across hospitals that actively submit data electronically versus those that do not (51.87%-54.14%), while actively submitting facilities demonstrate more consistent Medicare percentages (53.51%-54.13%). The SDs for Medicare percentages are generally higher in nonactive facilities (up to SD 20.15) compared to active facilities (up to SD 16.27). The HHI values for actively submitting facilities (ranging from 0.53 to 0.56, all with SD 0.36) are consistently lower than for nonactive facilities (ranging from 0.59 to 0.67, mostly with SD 0.37). Medicaid percentages are similar between active and nonactive facilities across all reporting categories, with active facilities showing slightly more consistent values (19.28%-20.02%) compared to nonactive facilities (18.6%-20.19%). SDs for Medicaid percentages are also generally higher in nonactive facilities.

XSL•FO
RenderX

Table . Hospital categorical characteristics across hospitals that actively submit data electronically versus those that do not.

Characteristics	Actively electronically submitting production data (yes or no), n (%)											
	Syndromic surveillance reporting		Immunization registry reporting		Electronic case reporting		Public health registry reporting		Clinical data registry reporting		Electronic reportable laboratory result reporting	
	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Ownership												
Government	59 (22.26)	236 (11.77)	36 (16.67)	261 (12.66)	184 (16.08)	111 (9.92)	148 (19.65)	135 (9.15)	170 (17.07)	121 (9.81)	74 (25.61)	219 (11.16)
For-profit	31 (11.7)	287 (14.31)	36 (16.67)	281 (13.63)	139 (12.15)	177 (15.82)	129 (17.13)	184 (12.47)	267 (26.81)	44 (3.57)	38 (13.15)	279 (14.21)
Not-for-profit	175 (66.04)	1482 (73.92)	144 (66.67)	1519 (73.7)	821 (71.77)	831 (74.26)	476 (63.21)	1157 (78.39)	559 (56.12)	1069 (86.63)	177 (61.25)	1465 (74.63)
Rural												
No	131 (49.43)	1295 (64.59)	104 (48.15)	1327 (64.39)	649 (56.73)	775 (69.26)	387 (51.39)	1010 (68.43)	574 (57.63)	828 (67.1)	134 (46.37)	1287 (65.56)
Yes	134 (50.57)	710 (35.41)	112 (51.85)	734 (35.61)	495 (43.27)	344 (30.74)	366 (48.61)	466 (31.57)	422 (42.37)	406 (32.9)	155 (53.63)	676 (34.44)
Size												
Small	167 (63.02)	887 (44.24)	135 (62.5)	923 (44.78)	585 (51.14)	463 (41.38)	437 (58.03)	600 (40.65)	517 (51.91)	512 (41.49)	190 (65.74)	850 (43.3)
Medium	81 (30.57)	819 (40.85)	66 (30.56)	836 (40.56)	425 (37.15)	475 (42.45)	245 (32.54)	637 (43.16)	373 (37.45)	514 (41.65)	78 (26.99)	818 (41.67)
Large	17 (6.42)	299 (14.91)	15 (6.94)	302 (14.65)	134 (11.71)	181 (16.18)	71 (9.43)	239 (16.19)	106 (10.64)	208 (16.86)	21 (7.27)	295 (15.03)
Part of a system												
No	99 (37.36)	377 (18.8)	61 (28.24)	418 (20.28)	303 (26.49)	171 (15.28)	231 (30.68)	238 (16.12)	283 (28.41)	184 (14.91)	115 (39.79)	355 (18.08)
Yes	166 (62.64)	1628 (81.2)	155 (71.76)	1643 (79.72)	841 (73.51)	948 (84.72)	522 (69.32)	1238 (83.88)	713 (71.59)	1050 (85.09)	174 (60.21)	1608 (81.92)
Teaching												
Not teaching	163 (61.51)	905 (45.14)	127 (58.8)	942 (45.71)	582 (50.87)	483 (43.16)	430 (57.1)	621 (42.07)	524 (52.61)	520 (42.14)	180 (62.28)	872 (44.42)
Minor	95 (35.85)	952 (47.48)	82 (37.96)	970 (47.06)	495 (43.27)	547 (48.88)	291 (38.65)	731 (49.53)	442 (44.38)	590 (47.81)	99 (34.26)	947 (48.24)
Major	7 (2.64)	148 (7.38)	7 (3.24)	149 (7.23)	67 (5.86)	89 (7.95)	32 (4.25)	124 (8.4)	30 (3.01)	124 (10.05)	10 (3.46)	144 (7.34)

XSL•FO
RenderX

Table . Hospital continuous characteristics across hospitals that actively submit data electronically versus those that do not.

Characteristics	Syndromic surveillance (n=2270), mean (SD)	Immunization registry (n=2277), mean (SD)	Electronic case (n=2263), mean (SD)	Public health registry (n=2229), mean (SD)	Clinical data registry (n=2230), mean (SD)	Electronic reportable laboratory results (n=2252), mean (SD)
Not actively electronically submitting production data						
Medicare Percentage	51.87 (20.15)	53.41 (18.56)	54.14 (17.19)	54 (17.99)	53.62 (16.69)	53.97 (18.98)
Medicaid Percentage	20.19 (15.14)	18.6 (12.84)	20.1 (14.23)	19.6 (14.89)	19.29 (14.01)	18.76 (14.66)
Herfindahl-Hirschman Index	0.62 (0.37)	0.62 (0.37)	0.61 (0.37)	0.64 (0.37)	0.59 (0.37)	0.67 (0.36)
Actively electronically submitting production data						
Medicare Percentage	54.13 (15.87)	53.89 (16.26)	53.51 (15.65)	53.9 (15.56)	54.02 (16.27)	53.88 (16.02)
Medicaid Percentage	19.59 (13.15)	19.77 (13.46)	19.28 (12.52)	19.57 (12.37)	20.02 (12.85)	19.84 (13.15)
Herfindahl-Hirschman Index	0.56 (0.36)	0.56 (0.36)	0.53 (0.36)	0.54 (0.36)	0.55 (0.36)	0.55 (0.36)
Total						
Medicare Percentage	53.86 (16.44)	53.85 (16.48)	53.83 (16.44)	53.94 (16.42)	53.84 (16.45)	53.89 (16.43)
Medicaid Percentage	19.66 (13.4)	19.66 (13.4)	19.69 (13.41)	19.58 (13.27)	19.69 (13.38)	19.7 (13.35)
Herfindahl-Hirschman Index	0.57 (0.36)	0.57 (0.36)	0.57 (0.36)	0.57 (0.36)	0.57 (0.36)	0.57 (0.36)

The statistical analysis using logistic regression models is shown in Table 3, which revealed several significant predictors of hospitals' engagement in electronic health data reporting across different reporting categories. For-profit hospitals show significantly lower odds of engaging in clinical data registry reporting compared to government hospitals (OR 0.15, 95% CI 0.09-0.22; $P<.001$), but higher odds for immunization registry reporting (OR 1.45, 95% CI 1.02-2.07; $P<.05$). Not-for-profit hospitals demonstrate significantly higher engagement in clinical data registry reporting (OR 1.89, 95% CI 1.43-2.50; $P<.001$), electronic case reporting (OR 1.76, 95% CI 1.25-2.48; $P<.01$), and public health registry reporting (OR 1.88, 95% CI 1.41-2.49; $P<.001$) compared to government-owned facilities.

Rural hospitals show significantly reduced likelihood of electronic reporting adoption across immunization registry (OR 0.77, 95% CI 0.61-0.97; $P<.05$), public health registry (OR 0.67, 95% CI 0.46-0.97; $P<.05$), and clinical data registry reporting (OR 0.77, 95% CI 0.60-0.98; $P<.05$) compared to urban counterparts. Hospital size emerges as a significant factor, with medium-sized hospitals showing higher engagement in electronic reportable laboratory results (OR 1.55, 95% CI 1.08-2.22; $P<.05$), public health registry (OR 1.51, 95% CI 1.02-2.25; $P<.05$), clinical data registry (OR 1.35, 95% CI 1.05-1.74; $P<.05$), and syndromic surveillance reporting (OR 1.52, 95% CI 1.06-2.19; $P<.05$) compared to small hospitals. Large hospitals demonstrate even stronger engagement in public health registry (OR 2.13, 95% CI 1.03-4.38; $P<.05$) and syndromic surveillance reporting (OR 2.29, 95% CI 1.17-4.46; $P<.05$).

System affiliation consistently emerges as one of the strongest predictors, with system-affiliated hospitals showing significantly higher odds of electronic reporting engagement across 5 of 6 categories: clinical data registry (OR 2.27, 95% CI 1.80-2.88; $P<.001$), immunization registry (OR 1.70, 95% CI 1.35-2.14; $P<.001$), electronic case reporting (OR 2.16, 95% CI 1.61-2.90; $P<.001$), public health registry (OR 1.78, 95% CI 1.42-2.25; $P<.001$), and electronic reportable laboratory results (OR 1.91, 95% CI 1.41-2.59; $P<.001$). Among teaching status variables, only major teaching hospitals show significantly higher odds for clinical data registry reporting (OR 2.66, 95% CI 1.56-4.52; $P<.001$). Medicare percentage shows a small but significant effect on syndromic surveillance reporting (OR 1.01, 95% CI 1.00-1.02; $P<.05$), while Medicaid percentage shows a minimal significant effect on immunization registry reporting (OR 0.99, 95% CI 0.98-1.00; $P<.05$). These small effect sizes for payer mix variables (ORs close to 1.0) suggest limited practical significance despite statistical significance, likely reflecting the large sample size rather than meaningful clinical impact.

XSL•FO
RenderX

Table . Logistic regression model of hospitals' engagement in electronic health data reporting across different reporting categories.

Characteristics	Clinical data registry reporting, OR[a] (95% CI)	Electronic case reporting, OR (95% CI)	Electronic reportable laboratory result, OR (95% CI)	Immunization registry reporting, OR (95% CI)	Public health registry reporting, OR (95% CI)	Syndromic surveillance reporting, OR (95% CI)
Ownership (reference: government)						
For-profit	0.15[b] (0.09-0.22)	1.45[c] (1.02-2.07)	1.38 (0.86-2.22)	0.77 (0.45-1.33)	0.95 (0.66-1.36)	1.45 (0.87-2.42)
Not-for-profit	1.89[b] (1.43-2.50)	1.31 (0.99-1.72)	1.76[d] (1.25-2.48)	1.11 (0.72-1.69)	1.88[b] (1.41-2.49)	1.42 (0.99-2.03)
Rural (reference: no)						
Yes	0.8 (0.62-1.02)	0.77[c] (0.61-0.97)	0.86 (0.62-1.19)	0.67[c] (0.46-0.97)	0.77[c] (0.60-0.98)	0.89 (0.63-1.25)
Size (reference: small)						
Medium	1.22 (0.95-1.58)	1.08 (0.85-1.37)	1.55[c] (1.08-2.22)	1.51[c] (1.02-2.25)	1.35[c] (1.05-1.74)	1.52[c] (1.06-2.19)
Large	1.12 (0.75-1.67)	1.19 (0.83-1.71)	1.85 (0.97-3.53)	2.13[c] (1.03-4.38)	1.43 (0.95-2.15)	2.29[c] (1.17-4.46)
Part of a system (reference: no)						
Yes	2.27[b] (1.80-2.88)	1.70[b] (1.35-2.14)	2.16[b] (1.61-2.90)	1.23 (0.87-1.75)	1.78[b] (1.42-2.25)	1.91[b] (1.41-2.59)
Teaching (reference: not teaching)						
Minor teaching	1.04 (0.82-1.30)	1.01 (0.82-1.25)	1.15 (0.84-1.58)	1.07 (0.76-1.52)	1.14 (0.91-1.43)	1.22 (0.88-1.68)
Major teaching	2.66[b] (1.56-4.52)	1.1 (0.71-1.70)	1.34 (0.58-3.11)	1.26 (0.49-3.28)	1.46 (0.86-2.46)	2.04 (0.81-5.16)
Medicare percentage	1.00 (1.00-1.01)	0.99[c] (0.98-1.00)	1.00 (0.99-1.01)	1.01 (1.00-1.02)	1.00 (0.99-1.01)	1.01[c] (1.00-1.02)
Medicaid percentage	1.00 (0.99-1.01)	0.99[d] (0.98-1.00)	1.00 (0.99-1.02)	1.01 (0.99-1.02)	0.99 (0.98-1.00)	1.00 (0.99-1.01)
Herfindahl-Hirschman Index	0.99 (0.72-1.36)	0.82 (0.61-1.10)	0.74 (0.48-1.14)	1.17 (0.73-1.88)	0.76 (0.56-1.04)	1.17 (0.76-1.81)

[a]OR: odds ratio.

[b]$P<.001$.

[c]$P<.05$.

[d]$P<.01$.

Automated Processes to Transmit Public Health Data

Table 4 reports hospital categorical characteristics across hospitals that have automated processes to transmit public health data (71.55% - 87.03% of "yes" responses), with particularly strong adoption for clinical data registry reporting (87.03%). Government hospitals show the lowest representation among automated reporting adopters (8.10% - 11.45%), while for-profit hospitals show moderate adoption that varies by reporting type, with notably higher representation in electronic case reporting (20.35%). The rural-urban divide is substantial, with nonrural hospitals constituting the clear majority of facilities using automated processes across all reporting categories (64.83% - 71.72%). The imbalance is most pronounced for clinical data registry reporting, where rural hospitals represent only 28.28% of automated adopters despite making up 40.86% of facilities not using automation for this purpose.

Hospital size shows a clear pattern where larger hospitals are disproportionately represented among automated process adopters. Medium and large hospitals together represent 55% to 60% of facilities using automation across reporting categories, despite making up only 40% to 47% of nonautomated facilities. Small hospitals, while still numerous among automation adopters (38.6% - 43.77%), are significantly under-represented compared to their share among nonautomated facilities (49.18% - 67.37%). System affiliation emerges as one of the strongest predictors, with system-affiliated hospitals representing 81.22% to 86% of facilities using automated processes across reporting categories. This contrasts sharply with their 62.84% to 76.13% representation among nonautomated facilities. Finally, teaching status also shows consistent patterns, with minor teaching and major teaching hospitals combined representing 51.25% to 60.54% of automated adopters across reporting categories, compared to 38.49% to 48.17% of nonautomated facilities. Major teaching hospitals, despite their small numbers overall, show consistently higher representation

XSL•FO
RenderX

among automated facilities (6.63% - 10.5%) compared to nonautomated ones (3.63% - 10.5%).

Table . Hospital categorical characteristics across hospitals that have automated processes to transmit public health data versus those that do not.

Characteristics	Automated processes to transmit the data (yes or no), n (%)											
	Syndromic surveillance reporting		Immunization registry reporting		Electronic case reporting		Public health registry reporting		Clinical data registry reporting		Electronic reportable laboratory result reporting	
	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Ownership												
Government	85 (20.38)	207 (11.27)	75 (22.66)	217 (11.45)	197 (18.11)	88 (8.1)	199 (18.27)	83 (8.41)	221 (15.84)	60 (8.75)	120 (23.35)	167 (10.07)
For-profit	36 (8.63)	281 (15.3)	51 (15.41)	245 (12.93)	72 (6.62)	221 (20.35)	69 (6.34)	158 (16.01)	198 (14.19)	29 (4.23)	54 (10.51)	197 (11.87)
Not-for-profit	296 (70.98)	1349 (73.43)	205 (61.93)	1433 (75.62)	819 (75.28)	777 (71.55)	821 (75.39)	746 (75.58)	976 (69.96)	597 (87.03)	340 (66.15)	1295 (78.06)
Rural												
No	228 (54.68)	1191 (64.83)	152 (45.92)	1237 (65.28)	611 (56.16)	752 (69.24)	627 (57.58)	686 (69.5)	825 (59.14)	492 (71.72)	265 (51.56)	1119 (67.45)
Yes	189 (45.32)	646 (35.17)	179 (54.08)	658 (34.72)	477 (43.84)	334 (30.76)	462 (42.42)	301 (30.5)	570 (40.86)	194 (28.28)	249 (48.44)	540 (32.55)
Size												
Small	240 (57.55)	804 (43.77)	223 (67.37)	819 (43.22)	582 (53.49)	429 (39.5)	579 (53.17)	381 (38.6)	686 (49.18)	275 (40.09)	302 (58.75)	700 (42.19)
Medium	133 (31.89)	765 (41.64)	83 (25.08)	788 (41.58)	355 (32.63)	500 (46.04)	363 (33.33)	447 (45.29)	515 (36.92)	300 (43.73)	153 (29.77)	706 (42.56)
Large	44 (10.55)	268 (14.59)	25 (7.55)	288 (15.2)	151 (13.88)	157 (14.46)	147 (13.5)	159 (16.11)	194 (13.91)	111 (16.18)	59 (11.48)	253 (15.25)
Part of a system												
No	125 (29.98)	345 (18.78)	123 (37.16)	351 (18.52)	301 (27.67)	152 (14)	299 (27.46)	154 (15.6)	333 (23.87)	117 (17.06)	171 (33.27)	295 (17.78)
Yes	292 (70.02)	1492 (81.22)	208 (62.84)	1544 (81.48)	787 (72.33)	934 (86)	790 (72.54)	833 (84.4)	1062 (76.13)	569 (82.94)	343 (66.73)	1364 (82.22)
Teaching												
Not teaching	237 (56.83)	822 (44.75)	211 (63.75)	834 (44.01)	564 (51.84)	454 (41.8)	562 (51.61)	401 (40.63)	690 (49.46)	277 (40.38)	295 (57.39)	720 (43.4)
Minor	155 (37.17)	885 (48.18)	108 (32.63)	918 (48.44)	444 (40.81)	560 (51.57)	450 (41.32)	511 (51.77)	626 (44.87)	337 (49.13)	184 (35.8)	821 (49.49)
Major	25 (6)	130 (7.08)	12 (3.63)	143 (7.55)	80 (7.35)	72 (6.63)	77 (7.07)	75 (7.6)	79 (5.66)	72 (10.5)	35 (6.81)	118 (7.11)

Table 5 reports hospital and market continuous characteristics across hospitals that have automated processes to transmit public health data versus those that do not.

The logistic regression analysis examined factors associated with hospitals' use of automated processes (EHR-generated data sent electronically or automatically) to transmit data to public health agencies across 6 reporting categories (Table 6). Hospital ownership was shown to significantly impact automated reporting practices. For-profit hospitals are 85% less likely than government hospitals to use automated processes for clinical data registry reporting (OR 0.15, 95% CI 0.09-0.22; $P<.001$), but 45% more likely to automate immunization registry reporting (OR 1.45, 95% CI 1.02-2.07; $P<.05$). Not-for-profit hospitals show significantly higher automation adoption in clinical data registry reporting (OR 1.89, 95% CI 1.43-2.50; $P<.001$), electronic case reporting (OR 1.76, 95% CI 1.25-2.48; $P<.01$), and public health registry reporting (OR 1.88, 95% CI 1.41-2.49; $P<.001$) compared to government facilities.

Rural status negatively impacts automation adoption, with rural hospitals showing significantly lower odds of automated data transmission for immunization registries (OR 0.77, 95% CI0.61-0.97; $P<.05$), public health registries (OR 0.67, 95% CI 0.46-0.97; $P<.05$), and clinical data registries (OR 0.77, 95% CI 0.60-0.98; $P<.05$). Hospital size matters, with medium-sized hospitals showing higher odds of automation across electronic reportable laboratory results (OR 1.55, 95% CI 1.08-2.22;

XSL•FO
RenderX

P<.05), immunization registries (OR 1.51, 95% CI 1.02-2.25; *P*<.05), public health registries (OR 1.35, 95% CI 1.05-1.74; *P*<.05), and syndromic surveillance (OR 1.52, 95% CI 1.06-2.19; *P*<.05) compared to small hospitals. Large hospitals show even stronger automation adoption in immunization registries (OR 2.13, 95% CI 1.03-4.38; *P*<.05) and syndromic surveillance (OR 2.29, 95% CI 1.17-4.46; *P*<.05).

System affiliation emerges as the most consistent predictor of automation adoption, with system-affiliated hospitals showing significantly higher odds of automated reporting across 5 categories: clinical data registry (OR 2.27, 95% CI 1.80-2.88; *P*<.001), immunization registry (OR 1.70, 95% CI 1.35-2.14; *P*<.001), electronic case reporting (OR 2.16, 95% CI 1.61-2.90; *P*<.001), public health registry (OR 1.78, 95% CI 1.42-2.25; *P*<.001), and electronic reportable laboratory results (OR 1.91, 95% CI 1.41-2.59; *P*<.001). Major teaching status significantly increases automation adoption for clinical data registry reporting (OR 2.66, 95% CI 1.56-4.52; *P*<.001), while Medicare and Medicaid percentages show minimal but significant effects on syndromic surveillance and immunization registry reporting, respectively. Market concentration (HHI) shows no significant association with automation adoption across all reporting categories.

Table . Hospital continuous characteristics across hospitals that have automated processes to transmit public health data versus those that do not.

Characteristics	Syndromic surveillance (n=2254), mean (SD)	Immunization registry (n=2226), mean (SD)	Electronic case (n=2174), mean (SD)	Public health registry (n=2076), mean (SD)	Clinical data registry (n=2081), mean (SD)	Electronic reportable laboratory results (n=2173), mean (SD)
No automated processes to transmit the data						
Medicare Percentage	52.69 (19.31)	54.51 (19.31)	53.79 (18.02)	53.77 (17.71)	53.41 (17.26)	53.95 (19.03)
Medicaid Percentage	20.29 (15.27)	18.24 (14.59)	19.69 (14.45)	19.56 (14.42)	19.87 (13.98)	19.1 (14.57)
Herfindahl-Hirschman Index	0.6 (0.36)	0.64 (0.35)	0.59 (0.36)	0.59 (0.36)	0.59 (0.36)	0.61 (0.36)
Automated processes to transmit the data						
Medicare Percentage	54.23 (15.69)	53.83 (16.04)	54.04 (14.89)	53.68 (15.24)	54.57 (15.28)	53.75 (15.69)
Medicaid Percentage	19.43 (12.85)	19.8 (13.15)	19.33 (12.1)	19.65 (12.04)	19.13 (12.08)	19.83 (12.96)
Herfindahl-Hirschman Index	0.56 (0.36)	0.56 (0.36)	0.56 (0.36)	0.53 (0.36)	0.52 (0.36)	0.55 (0.36)
Total						
Medicare Percentage	53.95 (16.43)	53.93 (16.56)	53.92 (16.52)	53.73 (16.58)	53.79 (16.64)	53.8 (16.54)
Medicaid Percentage	19.59 (13.33)	19.57 (13.39)	19.51 (13.33)	19.6 (13.34)	19.63 (13.39)	19.66 (13.36)
Herfindahl-Hirschman Index	0.57 (0.36)	0.58 (0.36)	0.57 (0.36)	0.56 (0.36)	0.57 (0.36)	0.56 (0.36)

XSL•FO
RenderX

Table . Logistic regression analysis of factors associated with hospitals' use of automated processes (electronic health record [EHR]-generated data sent electronically or automatically) to transmit data to public health agencies across 6 reporting categories.

Characteristics	Clinical data registry reporting, OR[a] (95% CI)	Electronic case reporting, OR (95% CI)	Electronic reportable laboratory result, OR (95% CI)	Immunization registry reporting, OR (95% CI)	Public health registry reporting, OR (95% CI)	Syndromic surveillance reporting, OR (95% CI)
Ownership (reference: government)						
For-profit	0.15[b] (0.09-0.22)	1.45[c] (1.02-2.07)	1.38 (0.86-2.22)	0.77 (0.45-1.33)	0.95 (0.66-1.36)	1.45 (0.87-2.42)
Not-for-profit	1.89[b] (1.43-2.50)	1.31 (0.99-1.72)	1.76[d] (1.25-2.48)	1.11 (0.72-1.69)	1.88[b] (1.41-2.49)	1.42 (0.99-2.03)
Rural (reference: no)						
Yes	0.8 (0.62-1.02)	0.77[c] (0.61-0.97)	0.86 (0.62-1.19)	0.67[c] (0.46-0.97)	0.77[c] (0.60-0.98)	0.89 (0.63-1.25)
Size (reference: small)						
Medium	1.22 (0.95-1.58)	1.08 (0.85-1.37)	1.55[c] (1.08-2.22)	1.51[c] (1.02-2.25)	1.35[c] (1.05-1.74)	1.52[c] (1.06-2.19)
Large	1.12 (0.75-1.67)	1.19 (0.83-1.71)	1.85 (0.97-3.53)	2.13[c] (1.03-4.38)	1.43 (0.95-2.15)	2.29[c] (1.17-4.46)
Part of a system (reference: no)						
Yes	2.27[b] (1.80-2.88)	1.70[b] (1.35-2.14)	2.16[b] (1.61-2.90)	1.23 (0.87-1.75)	1.78[b] (1.42-2.25)	1.91[b] (1.41-2.59)
Teaching (reference: not teaching)						
Minor teaching	1.04 (0.82-1.30)	1.01 (0.82-1.25)	1.15 (0.84-1.58)	1.07 (0.76-1.52)	1.14 (0.91-1.43)	1.22 (0.88-1.68)
Major teaching	2.66[b] (1.56-4.52)	1.1 (0.71-1.70)	1.34 (0.58-3.11)	1.26 (0.49-3.28)	1.46 (0.86-2.46)	2.04 (0.81-5.16)
Medicare Percentage	1.00 (1.00-1.01)	0.99[c] (0.98-1.00)	1.00 (0.99-1.01)	1.01 (1.00-1.02)	1.00 (0.99-1.01)	1.01[c] (1.00-1.02)
Medicaid Percentage	1.00 (0.99-1.01)	0.99[d] (0.98-1.00)	1.00 (0.99-1.02)	1.01 (0.99-1.02)	0.99 (0.98-1.00)	1.00 (0.99-1.01)
Herfindahl-Hirschman Index	0.99 (0.72-1.36)	0.82 (0.61-1.10)	0.74 (0.48-1.14)	1.17 (0.73-1.88)	0.76 (0.56-1.04)	1.17 (0.76-1.81)

[a]OR: odds ratio.

[b]$P<.001$.

[c]$P<.05$.

[d]$P<.01$.

Discussion

Principal Findings

This study identifies some of the main differences in automation and integrating public health information between US hospitals driven by structural resource inequalities, institutional practice, and location. Rural, independent, and smaller hospitals lag far behind urban, system-affiliated, and larger hospitals when it comes to adopting automated reporting systems. Despite national-level attempts to standardize health IT infrastructure, these gaps underscore systemic obstacles based on financial interests, organizational capacities, and market forces.

Rural hospitals continue to face significant challenges in adopting electronic public health reporting despite national progress in health IT adoption [15]. Limited financial resources and constrained operational capacity hinder their ability to invest in the infrastructure required for automation. These facilities often serve smaller patient populations and receive lower reimbursement rates, which makes it difficult to justify the high upfront costs of implementing advanced reporting systems. Additionally, rural hospitals typically lack access to IT specialists and foundational systems that support seamless electronic data exchange, resulting in a greater reliance on manual or mixed reporting methods. These barriers not only restrict their compliance with public health reporting requirements but also widen the digital divide between rural and urban health care providers. Addressing these disparities requires targeted policy support and financial investment to ensure rural hospitals can fully participate in the public health data ecosystem.

XSL•FO
RenderX

The nonrural versus rural divide is stark in the results as both the rates of actively submitting data electronically and the adoption of automated processes to transmit that public health data show low rates of submission and adoption by rural hospitals in all reporting categories. There are many possible reasons for this difference largely relating to the differing economic environments of these hospitals. Rural hospitals often face greater financial strain due to the poorer socioeconomic conditions of their locals and thus do not have the financial capital to invest in high-tech systems. As Younis [17] shared that rural hospitals generate less revenue than urban hospitals and are significantly disadvantaged in terms of performance.

Another avenue to look at is the role of competition from other hospitals that nonrural hospitals face. As discussed in Garcia-Lacalle and Martin [18], hospitals in a market-driven environment have a keen sense of where they sit in comparison to their competition and therefore consider new strategies to better focus on patients and users. Once one hospital in a competitive environment adopts an electronic data submission system or automates their pre-existing one, it encourages other hospitals in that same environment to also adopt. Similarly, Ghiasi et al [19] found, in their literature review, that hospitals in a competitive market seek to differentiate themselves from competitors through specific services. Some of these differentiating services could be electronic data submission systems.

In our study, larger hospitals benefit from centralized IT infrastructure and specialized personnel, enabling consistent compliance with evolving standards. Medium and large hospitals show 1.5 to 2.3 times higher odds of automation across categories like syndromic surveillance and laboratory reporting. These institutions absorb upfront costs more effectively and maintain robust EHR systems, whereas smaller facilities struggle with limited staffing and budgetary flexibility. Particularly larger hospitals within multihospital systems demonstrate higher engagement in both active electronic data submission and automated reporting due to greater resource availability. These hospitals benefit from economies of scale that support investment in centralized IT infrastructures and EHR systems. In addition, system-affiliated hospitals are also more likely to have internal health IT teams and established workflows for public health communication because it reduces barriers to implementation.

Not-for-profit hospitals lead in adoption due to mission-driven commitments to population health and access to grant funding. Their focus on community benefit aligns with public health reporting goals, whereas for-profit hospitals prioritize revenue-generating technologies (eg, billing systems). Government hospitals, with the limitations of bureaucratic procurement systems, fall behind despite regulatory encouragement. The trends are indicative of findings by Tsai et al [20] that financial restrictions and fragmented workflows are the main barriers against EHR adoption in low-resource settings.

In our study, facilities not actively submitting data electronically exhibit more variable Medicare percentages (51.87% - 54.14%), suggesting that markets with less competition (higher HHI values) reduce pressure to adopt reporting technologies. Lower digital literacy among older Medicare populations may also deprioritize automation in regions serving these demographics. Conversely, hospitals in competitive, high-volume markets align IT investments with performance metrics to meet patient and regulatory expectations.

Policy Implications

A 2024 analysis by the Kaiser Family Foundation found that nearly half of US metropolitan areas are dominated by just one or two hospital systems, significantly reducing competition and, consequently, the urgency for these institutions to adopt advanced data reporting practices [21]. This aligns with findings from the BMC Health Services Research, which revealed that providers in rural or less competitive regions demonstrate lower EHR adoption and interoperability [22]. Moreover, patient population characteristics, particularly among older adults on Medicare, further influence reporting engagement. A systematic review in the Archives of Public Health emphasized the digital health literacy gap in this group, suggesting that facilities serving older or underserved populations may deprioritize electronic data initiatives due to lower patient engagement with digital platforms [23]. These studies underscore the multifactorial barriers to robust public health data reporting, suggesting the need for targeted policy incentives and infrastructure support to promote broader and more equitable adoption.

Limitations

This study's limitation lies in its reliance on secondary data from the 2023 AHA Annual Survey. Hospital characteristics are based on self-reported data which may affect accuracy. Our cross-sectional design limits causal inference. This analysis focused on US hospitals only, affecting generalizability to other types of health organizations and countries. Finally, the rural or urban classification using Rural-Urban Commuting Area codes may not fully capture rural-urban distinctions. Residual confounding may exist due to unmeasured variables such as IT staffing levels or leadership engagement.

Conclusions

The clear difference between nonrural and rural hospitals in terms of electronic data submission and automation adoption shows significant gaps caused by economic and competitive factors. Nonrural hospitals, benefiting from higher revenue and competitive pressures, are more likely to invest in advanced IT systems and automated processes. On the other hand, rural hospitals face financial constraints and lower patient volumes, limiting their ability to adopt such technologies. This divide is further worsened by the centralized resource allocation and organized workflows in system-aligned hospitals, which improve their reporting capabilities. Not-for-profit hospitals also lead in electronic health data adoption due to their mission-driven priorities and access to grant funding. Research highlights the many barriers to strong public health data reporting, shaped by market dynamics and patient demographics. Effective strategies for improving electronic data submission may include tailored incentives, strategic partnerships, and population-specific approaches. Addressing these gaps is crucial for ensuring fair access to advanced health care technologies and improving

XSL•FO
RenderX

overall public health reporting. Targeted policy interventions and financial support are essential to help rural hospitals overcome structural barriers and participate more fully in the nation's public health data system.

Data Availability

Data used in this study are available from the American Hospital Association and are available for purchase through a data usage agreement.

Authors' Contributions

HH conceived and designed the study, including the research questions, methodology, and analytical approach; supervised all aspects of the research and data analysis; interpreted the results and developed the initial findings; provided critical revision of the manuscript for intellectual content; ensured all aspects of the research were accurately reflected; and took final responsibility for submission.

AA contributed to study design and analytical approach; wrote sections of the manuscript; and participated in revision and editing of the final version.

CSL assisted in developing the study design and analytical framework; participated in data analysis and interpretation; wrote sections of the manuscript; and contributed to manuscript revision for accuracy and clarity.

DB participated in data management and preparation; assisted with statistical analysis; wrote sections of the manuscript; and provided critical feedback on analytical approaches.

CSL assisted in developing the study design and analytical framework; participated in data analysis and interpretation; wrote sections of the manuscript; and contributed to manuscript revision for accuracy and clarity.

JYJ participated in data analysis and visualization; assisted with results interpretation; contributed to manuscript drafting; and provided critical feedback during revision.

Conflicts of Interest

None declared.

References

1. de Groot R, van der Graaff F, van der Doelen D, et al. Implementing findable, accessible, interoperable, reusable (FAIR) principles in child and adolescent mental health research: mixed methods approach. JMIR Ment Health 2024 Dec 19;11:e59113. [doi: 10.2196/59113] [Medline: 39727091]
2. Inau ET, Sack J, Waltemath D, Zeleke AA. Initiatives, concepts, and implementation practices of the findable, accessible, interoperable, and reusable data principles in health data stewardship: scoping review. J Med Internet Res 2023 Aug 28;25:e45013. [doi: 10.2196/45013] [Medline: 37639292]
3. American Recovery and Reinvestment Act of 2009. : Congress.gov; 2009 URL: https://www.congress.gov/bill/111th-congress/house-bill/1/text/pl [accessed 2026-01-16]
4. Annual report fiscal year 2009. : Library of Congress; 2009 URL: https://www.loc.gov/crsinfo/about/crs09_annrpt.pdf [accessed 2026-01-16]
5. Zayas-Cabán T, Haque SN, Kemper N. Identifying opportunities for workflow automation in health care: lessons learned from other industries. Appl Clin Inform 2021 May;12(3):686-697. [doi: 10.1055/s-0041-1731744] [Medline: 34320683]
6. Alam MA, et al. Real-time analytics in streaming big data: techniques and applications. J Sci Eng Res 2024;1(1):104-122. [doi: 10.70008/jeser.v1i01.56]
7. Ugajin A. Automation in hospitals and health care. In: Nof SY, editor. Springer Handbook of Automation: Springer; 2023:1209-1233. [doi: 10.1007/978-3-030-96729-1_56]
8. Public reporting mission statement. American College of Cardiology – Quality Improvement for Institutions. 2024. URL: https://cvquality.acc.org/NCDR-Home/registries/acc-public-reporting/public-reporting-mission-statement [accessed 2026-01-16]
9. ACC public reporting. American College of Cardiology – Quality Improvement for Institutions. 2024. URL: https://cvquality.acc.org/NCDR-Home/registries/acc-public-reporting [accessed 2026-01-16]
10. Jiang JX, Cram P, Qi K, Bai G. Challenges and dynamics of public health reporting and data exchange during COVID-19: insights from US hospitals. Health Aff Sch 2024 Jan;2(1):qxad080. [doi: 10.1093/haschl/qxad080] [Medline: 38756405]
11. Alubaie MA, Sayed MY, Alnakhli RE, et al. The efficiency and accuracy gains of real-time health data integration in healthcare management: a comprehensive review of current practices and future directions. Egypt J Chem 2024 Dec 1;67(13):1725-1729. [doi: 10.21608/ejchem.2025.343595.10967]
12. Poppe O, Sæbø JI, Braa J. Strategies for standardizing health information analysis. 2019 Presented at: 15th International Conference on Social Implications of Computers in Developing Countries (ICT4D 2019); May 20-24, 2019. [doi: 10.1007/978-3-030-18400-1_21]

XSL•FO
RenderX

13. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: a more personalized healthcare and public health policy. Health Policy Technol 2017 Mar;6(1):20-25. [doi: 10.1016/j.hlpt.2016.08.002]
14. 2023 AHA annual survey. American Hospital Association. 2023 URL: https://www.ahadata.com/system/files/media/file/2024/09/AHA-Annual-Survey-2023.pdf [accessed 2026-01-16]
15. Richwine C. Progress and ongoing challenges to electronic public health reporting among non-federal acute care hospitals. : Office of National Coordinator For Health Information Technology; 2023 URL: https://pubmed.ncbi.nlm.nih.gov/39475518/ [accessed 2026-01-16] [Medline: 39475518]
16. Lu L, Pan J. Does hospital competition lead to medical equipment expansion? Evidence on the medical arms race. Health Care Manag Sci 2021 Sep;24(3):582-596. [doi: 10.1007/s10729-020-09529-x] [Medline: 33411086]
17. Younis MZ. A comparison study of urban and small rural hospitals financial and economic performance. Online J Rural Nurs Health Care 2003;3(1):38-48. [doi: 10.14574/ojrnhc.v3i1.247]
18. Garcia-Lacalle J, Martin E. Rural vs urban hospital performance in a “competitive” public health service. Soc Sci Med 2010 Sep;71(6):1131-1140. [doi: 10.1016/j.socscimed.2010.05.043] [Medline: 20630645]
19. Ghiasi A, et al. The impact of hospital competition on strategies and outcomes of hospitals: a systematic review of the U.S. hospitals 1996-2016. J Health Care Finance 2018;44(2) [FREE Full text]
20. Tsai CH, Eghdam A, Davoody N, Wright G, Flowerday S, Koch S. Effects of electronic health record implementation and barriers to adoption and use: a scoping review and qualitative analysis of the content. Life (Basel) 2020 Dec 4;10(12):327. [doi: 10.3390/life10120327] [Medline: 33291615]
21. Levinson Z, et al. Key facts about hospitals. Kaiser Family Foundation. 2025. URL: https://www.kff.org/health-costs/key-facts-about-hospitals/?entry=overview-introduction [accessed 2026-01-16]
22. Anzalone AJ, Geary CR, Dai R, Watanabe-Galloway S, McClay JC, Campbell JR. Lower electronic health record adoption and interoperability in rural versus urban physician participants: a cross-sectional analysis from the CMS quality payment program. BMC Health Serv Res 2025 Jan 23;25(1):128. [doi: 10.1186/s12913-024-12168-5] [Medline: 39849475]
23. Mumtaz H, Riaz MH, Wajid H, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023;5:1203945. [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]

Abbreviations

AHA: American Hospital Association
AI: artificial intelligence
EHR: electronic health record
HHI: Herfindahl-Hirschman Index
OR: odds ratio

Edited by E Mensah; submitted 21.Oct.2025; peer-reviewed by NMA Zakari, S Mirji; revised version received 12.Dec.2025; accepted 22.Dec.2025; published 06.Feb.2026.

Please cite as:
Hamadi H, Alford A, Smith Lopez C, Baker D, Samaniego JI, Jin JY
Associations Between Hospital Structural Characteristics and Adoption of Public Health Data Integration and Automation: National Cross-Sectional proofsStudy
Online J Public Health Inform 2026;18:e86263
URL: https://ojphi.jmir.org/2026/1/e86263
doi:10.2196/86263

© Hanadi Hamadi, Alaysia Alford, Chloe Smith Lopez, Durron Baker, James Ian Samaniego, Jessica Yu Jin. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 6.Feb.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Associations Between Digital Attitudes, Digital Inequalities, and Health Care Use Among Recipients of a Community-Based Laptop Distribution Program: Cross-Sectional Study

Tianca Crocker, PhD

College of Health and Human Services, School of Social Work, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC, United States

Corresponding Author:
Tianca Crocker, PhD
College of Health and Human Services, School of Social Work, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC, United States

Abstract

Background: Telehealth use expanded rapidly during the COVID-19 pandemic, yet persistent digital inequities continue to shape who benefits from virtual and preventive health services. Although community-based digital inclusion programs aim to reduce barriers to access, less is known about how device reliability, broadband conditions, and perceptions of online health information influence health care use after access is provided.

Objective: The objective of this study was to investigate associations between digital attitudes and access and health care use, including telehealth use and preventive health care visits, among participants in a community-based laptop distribution program.

Methods: This study analyzed cross-sectional survey data from 1019 participants in a community-based laptop distribution program. Refurbished laptops were distributed between October 2020 and July 2023, after which approximately 20,000 program recipients were invited by email to complete an online survey in English or Spanish. Ordered logistic regression models assessed predictors of telehealth use and preventive health care use. Models were adjusted for sociodemographic characteristics, digital access conditions, and attitudes toward online health information.

Results: Among respondents, 26.95% (218/809) reported increased preventive health care visits since receiving a laptop, and 62% (502/810) reported using telehealth sometimes or often. Having a functioning program laptop was associated with higher odds of increased preventive health care use (odds ratio [OR] 2.13, 95% CI 1.07 - 4.27; *P*=.03), whereas living with a partner was associated with lower odds (OR 0.47, 95% CI 0.23 - 0.99; *P*=.048). Participants with both fast internet and affordable broadband had greater odds of increased preventive health care use (OR 8.29, 95% CI 1.03 - 66.50; *P*=.046). Participants who strongly agreed that online information helped them make better health decisions were more likely to report increased preventive health care use (OR 15.03, 95% CI 4.53 - 49.82; *P*<.001). More frequent telehealth use was associated with reporting that online health information often improved health (OR 17.77, 95% CI 6.78 - 46.56; *P*<.001) and with speaking a language other than English as a primary language (OR 5.95, 95% CI 1.73 - 20.41; *P*=.005). Respondents reporting decreased physician visits had lower odds of telehealth use (OR 0.21, 95% CI 0.10 - 0.46; *P*<.001). Race, ethnicity, gender, race-by-gender, and ethnicity-by-gender interaction terms were not substantial in adjusted models.

Conclusions: Digital inclusion efforts may support engagement with preventive and virtual health services, but access alone appears insufficient. Device functionality, broadband quality and affordability, and perceptions of the value of online health information were important predictors of health care use. Digital inclusion initiatives should address both technological infrastructure and digital health engagement to promote equitable access to care.

(Online J Public Health Inform 2026;18:e97872) doi:10.2196/97872

KEYWORDS

health care use; computers; libraries; internet-based intervention; digital health; health knowledge; attitudes; practice

Introduction

Telehealth use in the United States grew rapidly during the COVID-19 pandemic as health care systems sought to deliver care amid social distancing, rising emergency service demand, and shifting regulations [1]. Despite this expansion, long-standing digital inequities persisted [2]. Individuals with low digital literacy, racial and ethnic minority groups, adults with limited English proficiency, and those lacking reliable or affordable broadband remained less likely to benefit from digital care [3-6]. Prior studies identify digital divides—limited digital literacy, devices, and/or broadband—as key determinants of

XSL•FO
RenderX

health and telehealth adoption [1,7,8], yet less is known about the frequency of telehealth use and preventive health care visits amid multifaceted digital barriers.

Digital divides operate at multiple levels: first-level divides involve disparities in internet access, digital devices, and basic navigation; second-level divides involve digital skills and use typologies; and third-level divides shape long-term well-being [4]. Research rarely addresses intersectional divides or examines how variations in first-level factors—such as device type or broadband cost—create barriers among those technically "connected" [1-12]. Intersecting identities—race, gender, and language—further compound these divides and may uniquely affect telehealth and preventive care use [10]. Black and Hispanic women sit at the intersection of structural inequities, exclusion, and digital skill gaps in the United States, yet limited research evaluates how these layered disadvantages influence health behaviors [13]. This paper presents preliminary health-related findings from a digital inclusion program that distributed free laptops during the pandemic. The aim of this study was to investigate the impact of digital divides on telehealth and preventive care use.

Methods

Overview

A cross-sectional design was used to analyze the study convenience sample (n=1019) drawn from 20,535 program participants who were eligible for the program based on residency in a large urban county in the southeastern United States. The survey was developed with input from the program's evaluation team and refined through multiple reviews to ensure clarity, appropriate length, and relevance to evaluation objectives. Email addresses of program participants who received a laptop were provided to evaluators for the sole purpose of distributing the online participant survey. Public health officials from the study's focal county reported rising or persistent health needs (ie, infant mortality, chronic disease, homicide, and accidents) [11] during this time. Respondents completed a 26-item survey capturing sociodemographic characteristics, digital attitudes and resources, and changes in well-being behaviors since receiving the free laptop. Telehealth use was measured as an ordered categorical outcome reflecting virtual visit frequency, and preventive health care use captured changes in the frequency of preventive health visits since laptop receipt. Key digital variables included digital literacy, broadband affordability and speed, number of household devices, and laptop functionality.

Two measures assessed distinct aspects of online health information use. The item "I have made better decisions about my health or medical care as a result of the information/advice I found online" measured respondents' perceived ability to use online health information in health-related decision-making. The item "How often have you found online information that helped you improve your health?" measured the frequency with which respondents perceived obtaining beneficial health outcomes from online information. Although conceptually related, these measures were intended to capture different dimensions of online health information engagement. To assess the degree of overlap between these measures, a Spearman correlation was calculated. The correlation was modest (ρ=0.32; P<.001), indicating that while the constructs were related, they captured distinct aspects of respondents' experiences with online health information. Additional variables captured well-being outcomes and lifestyle changes due to online information seeking. Models were adjusted for employment, primary language, marital status, Black women, Hispanic women, age, and education. Ordered logistic regression estimated predictors of the frequency of telehealth use and preventive health care visits following device receipt. Statistical analyses were performed in Stata (version 18; StataCorp).

Ethical Considerations

Survey responses were captured electronically using Qualtrics and stored in a secure, access-restricted Dropbox repository to protect participant confidentiality and data integrity. Survey respondents received a US $5 incentive that was delivered electronically following survey completion. Institutional review board (IRB-22-0707) approval was obtained from the University of North Carolina at Charlotte's Office of Research Protections and Integrity. All participants provided informed consent before participating in this study and all data were deidentified prior to analysis to protect participant privacy and maintain confidentiality.

Results

This study found that digital access and perceived benefits of online health information were significantly associated with both preventive health care use and telehealth use. Participants with a functioning program laptop and affordable, high-speed internet were more likely to report increased preventive health care visits, while stronger beliefs that online health information improved health were consistently associated with greater use of both preventive care and telehealth. A total of 1019 program participants completed the survey, with sample sizes varying by variable due to the optional item-response design of the online survey. Respondents had a mean age of 41.39 (SD 28.54) years. Of 761 respondents, 528 (69.38%) identified as Black or African American, and 85 (11.17%) identified as Hispanic. More than half (446/818, 54.52%) of the respondents worked full time, and almost all (790/834, 94.72%) reported English as their primary language. Approximately one-quarter (218/809, 26.95%) of respondents reported increased preventive health care visits after receiving a laptop, and 61.98% (502/810) of respondents reported general telehealth use as "often" or "sometimes" on a 5-point Likert scale. Additional descriptives are reported in Table S1 in Multimedia Appendix 1.

Two ordered logistic regression models assessed predictors of increased preventive health care visits since receipt of the laptop (model A) and frequency of telehealth visits (model B). Refer to Table S2 in Multimedia Appendix 1. Model A demonstrated a strong fit (likelihood ratio χ^2_{44}=82.3; P<.001; pseudo R^2=0.1029). Model B also showed strong fit (likelihood ratio χ^2_{43}=180.2; P<.001; pseudo R^2=0.1389).

Model A showed that living with a partner reduced the odds of increased preventive visits (odds ratio [OR] 0.47, 95% CI

XSL•FO
RenderX

0.23-0.99; *P*=.048). Neutral reports of lifestyle confidence gained from online information predicted lower odds (OR 0.26, 95% CI 0.07-0.97; *P*=.04) of visits. Of the respondents, 26.95% (218/809) increased preventive health care use since receiving the device, and respondents with a functioning program laptop had more than twice the odds of increased visits (OR 2.13, 95% CI 1.07-4.27; *P*=.03). However, a functioning laptop interacted with having enough digital devices in the home, suggesting diminished program benefits for health care uptake among laptop recipients who also had access to other digital devices in the home (OR 0.35, 95% CI 0.14-0.91; *P*=.03). In addition, a significant interaction between having a fast internet connection and affordable home broadband was observed, indicating that the combined presence of speed and affordability was associated with substantially higher odds of increased preventive health care use (OR 8.29, 95% CI 1.03-66.50; *P*=.046). Beliefs about the value of online health information were strongly associated with increased preventive health care use following laptop receipt. Relative to respondents who strongly disagreed that online information helped them make better health decisions, those who somewhat disagreed had significantly higher odds of reporting increased preventive health care use (OR 5.51, 95% CI 1.52 - 19.92; *P*=.009). The odds were even higher among respondents who neither agreed nor disagreed (OR 11.76, 95% CI 3.61 - 38.30; *P*<.001), somewhat agreed (OR 18.27, 95% CI 5.88 - 56.83; *P*<.001), and strongly agreed (OR 15.03, 95% CI 4.53 - 49.82; *P*<.001), suggesting a positive relationship between perceived usefulness of online health information and increased preventive health care use after receiving a laptop.

Model B showed that divorced respondents more frequently used telehealth (OR 2.04, 95% CI 1.08-3.87; *P*=.03). Fewer physician visits predicted substantially less frequent use of telehealth (OR 0.21, 95% CI 0.10-0.46; *P*<.001). Strong beliefs among survey participants that their health benefited from online health information predicted more frequent telehealth use (ORs 4.37 [95% CI 1.67-11.47] to 17.77 [95% CI 6.78-46.56]). A functioning laptop predicted less frequent reported telehealth use (OR 0.52, 95% CI 0.29-0.94; *P*=.03). Non-English speakers reported higher telehealth use (OR 5.95, 95% CI 1.73-20.41; *P*=.01). Generalized ordered logit models were explored to relax proportional odds assumptions; however, the telehealth model did not converge, suggesting sparse subgroup patterns.

Discussion

Principal Findings

The results suggest that digital inclusion initiatives may support health care engagement, but that their benefits vary according to broader digital access and resource contexts. Furthermore, these findings draw attention to layers within first- and second-order digital inequities that shape health care use while simultaneously offering deeper context for public health officials and practitioners seeking metrics to explore digital structural barriers in urban environments. Patterns of health engagement suggest that digital access alone is insufficient; instead, meaningful online health information experiences; device functioning; and fast, affordable home broadband play an important role in how recipients use both in-person and virtual health care.

Model A

While unexpected, the significant finding that participants living with a partner were less likely to increase preventive visits after receiving the device suggests that household dynamics may influence the uptake of new digital resources, and further exploration of marital status is warranted. Digital navigators may benefit from tailoring support at the household level, especially for individuals who rely on shared devices or caregiving arrangements that constrain health-seeking activities.

A functioning program laptop was associated with increased preventive care visits, underscoring the importance of device reliability. Digital community programs may consider extending repair or technical support to sustain long-term user health engagement. Although greater digital literacy did not consistently predict more frequent telehealth use, an important finding was the strong association between perceptions of online health information and increased preventive health care use after laptop receipt. Even respondents who only somewhat disagreed that online information improved their health decision-making had significantly greater odds of reporting increased preventive care use relative to those who strongly disagreed. This pattern suggests that complete skepticism toward the health value of online information may be associated with lower engagement in preventive care, whereas even modest openness to using online information may facilitate health care engagement. Although causality cannot be inferred, the findings are consistent with the possibility that individuals who view online health information as useful may be more likely to seek preventive services when digital access barriers are reduced. Additionally, the observed interaction between fast and affordable broadband underscores the importance of considering digital infrastructure as a social determinant of health rather than as a solitary first-order access metric. Preventive care use increased most clearly among participants who reported both adequate connection speed and affordable service, suggesting that broadband speed alone or affordability alone may be insufficient to support sustained engagement with health care services. These findings highlight how infrastructure quality may shape digital divides that translate digital access into measurable health behaviors.

Model B

Model B also highlighted the importance of digital access conditions in shaping participants' online behavior. Telehealth use varied across sociodemographic contexts. More frequent telehealth use among divorced respondents and non-English speakers may indicate an opportunity for culturally responsive and linguistically tailored support. Despite the statistical significance of the association, the small proportion of non–English-speaking respondents suggests that replication in more linguistically diverse samples is needed before drawing broader conclusions. Conversely, participants reporting fewer physician visits after receipt of the laptop were less likely to use telehealth, suggesting that patient-health care provider trust continues to shape digital health care engagement.

XSL•FO
RenderX

In addition to sociodemographic and experiential factors, telehealth use was associated with device and connectivity conditions. Participants reporting that the program laptop was still functioning used telehealth less frequently, while those reporting that online information improved their health used telehealth substantially more frequently. These patterns suggest that telehealth engagement may decline as digital access stabilizes, whereas perceived usefulness of online health information could be a key driver of sustained virtual care use; however, causal inferences cannot be drawn from the observed associations. Additionally, the negative association between having a functioning laptop and frequency of telehealth use was somewhat counterintuitive. However, unlike model A, which focused on preventive health care use since receipt of the program laptop, model B assessed broader telehealth use that may have occurred through multiple technology platforms or prior to receipt of the laptop. As smartphone-based telehealth use has increased [14], participants may have relied on mobile devices rather than laptops to access virtual care. This finding may also reflect unmeasured differences in telehealth preferences or digital engagement and should therefore be considered exploratory pending further investigation.

The relatively low model percentage of variance, or R^2 values, suggests that a substantial proportion of variability in health care use remains unexplained. Health care–seeking behavior is influenced by numerous factors beyond digital access and online health information use, including health status, chronic disease burden, insurance coverage, health care provider availability, transportation access, health literacy, and digital navigation support. Future studies incorporating these factors may provide a more comprehensive understanding of telehealth and preventive care use.

Limitations

The study relied on cross-sectional, single-item, self-reported data and could not establish causal relationships. The analytic sample was drawn from participants in a single urban digital inclusion initiative, which may limit generalizability to other settings or populations. In addition, longitudinal patterns of telehealth use following device receipt could not be assessed. Despite these limitations, the findings provide timely, practice-relevant evidence for digital inclusion strategies that extend beyond device distribution to address connectivity quality, affordability, and health-relevant digital engagement.

Conclusions

This study demonstrated that digital inequities may shape telehealth use and preventive health care engagement even within a community-based digital inclusion program. Findings indicate that device reliability, broadband affordability and speed, and perceived usefulness of online health information are more strongly associated with health service use than access alone. Improving digital access may be an important consideration for health care engagement; however, future longitudinal and experimental studies are needed to determine whether these factors directly influence health care use. The results call attention to digital infrastructure and broader information environments as actionable areas to explore the social determinants of preventive care and virtual health engagement.

Funding

The author declared no financial support was received for this work.

Data Availability

The data supporting the findings of this study are not publicly available.

Authors' Contributions

This study is solely the work of the author, and no external reviewers contributed to the development of the manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive and inferential statistics tables.
[DOCX File, 20 KB - ojphi_v18i1e97872_app1.docx]

References

1. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022 Dec;49(4):517-530. [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
2. Beckett MK, Haas A, Saliba D, et al. Gaps in internet use narrowed among older adults with Medicare during the COVID-19 pandemic but persist. J Am Geriatr Soc 2024 Apr;72(4):1283-1287. [doi: 10.1111/jgs.18735] [Medline: 38190414]
3. Eberly LA, Kallan MJ, Julien HM, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 1;3(12):e2031640. [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]

XSL•FO
RenderX

4. Goldstein KM, Zullig LL, Dedert EA, et al. Telehealth interventions designed for women: an evidence map. J Gen Intern Med 2018 Dec;33(12):2191-2200. [doi: 10.1007/s11606-018-4655-8] [Medline: 30284173]
5. Husain L, Greenhalgh T, Hughes G, Finlay T, Wherton J. Desperately seeking intersectionality in digital health disparity research: narrative review to inform a richer theorization of multiple disadvantage. J Med Internet Res 2022 Dec 7;24(12):e42358. [doi: 10.2196/42358] [Medline: 36383632]
6. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 1;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
7. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the COVID-19 pandemic. NEJM Catal Innov Care Deliv 2020;1:1-13. [doi: 10.1056/CAT.20.0123]
8. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52. [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
9. Acilar A, Sæbø Ø. Towards understanding the gender digital divide: a systematic literature review. Glob Knowl Mem Commun 2023 Feb 20;72(3):233-249. [doi: 10.1108/GKMC-09-2021-0147]
10. Early J, Hernandez A. Digital disenfranchisement and COVID-19: broadband internet access as a social determinant of health. Health Promot Pract 2021 Sep;22(5):605-610. [doi: 10.1177/15248399211014490] [Medline: 33955266]
11. Connecting communities: Mecklenburg County 2022 community health assessment. Mecklenburg County North Carolina. URL: https://mecknc.widen.net/s/pqddccfcsn/mc-doc-hlt-cha-2022 [accessed 2026-02-03]
12. Medero K, Merrill K Jr, Ross MQ. Modeling access across the digital divide for intersectional groups seeking web-based health information: national survey. J Med Internet Res 2022 Mar 15;24(3):e32678. [doi: 10.2196/32678] [Medline: 35289761]
13. Franco K, Patler C, Pirtle WL. COVID-19's unequal toll: differences in health-related quality of life by gendered and racialized groups. J Health Soc Behav 2024 Mar;65(1):60-74. [doi: 10.1177/00221465231199734] [Medline: 37830769]
14. Bolbocean C, Hayes C, Bogulski C. Telehealth use and modality choice among US adults: Shorrocks-Shapley decomposition of a 2022 cross-sectional national survey. J Med Internet Res 2026 Mar 18;28:e81879. [doi: 10.2196/81879] [Medline: 41849671]

Abbreviations

OR: odds ratio

Edited by E Mensah; submitted 10.Apr.2026; peer-reviewed by Z Liu; revised version received 03.Jul.2026; accepted 12.Jul.2026; published 06.Oct.2026.

Please cite as:
Crocker T
Associations Between Digital Attitudes, Digital Inequalities, and Health Care Use Among Recipients of a Community-Based Laptop Distribution Program: Cross-Sectional Study
Online J Public Health Inform 2026;18:e97872
URL: https://ojphi.jmir.org/2026/1/e97872
doi:10.2196/97872

© Tianca Crocker. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 6.Oct.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Financial Literacy and Financial Toxicity Among US Veterans: Cross-Sectional Survey Informing Public Health Informatics Screening

Harman Sabharwal[1]; Ramona Raya[2], MD

[1]Langley High School, 6520 Georgetown Pike, McLean, VA, United States

[2]Department of Medical Education, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Harman Sabharwal
Langley High School, 6520 Georgetown Pike, McLean, VA, United States

Abstract

Background: Financial toxicity can contribute to adverse health and care-access outcomes among US veterans, yet scalable methods to identify individuals at elevated risk remain limited. Public health informatics frameworks may enable the translation of patient-reported financial risk signals into streamlined screening, risk stratification, and care-navigation workflows.

Objective: This study aimed to examine concept-level indicators of financial literacy and financial toxicity among US veterans and explore how these findings could inform future informatics-enabled screening strategies for identifying subgroups at increased risk of health-related financial strain.

Methods: We conducted an exploratory cross-sectional survey of 88 US veterans from 2024 to 2025. Financial literacy was assessed using 3 benchmark items from the National Financial Capability Study. Financial toxicity was assessed using items aligned with domains reflected in the Comprehensive Score for Financial Toxicity framework, including difficulty affording care, reduced or quit work, borrowing money or using savings for care, and treatment-adherence impact. Analyses included descriptive statistics, Fisher exact tests, unadjusted logistic regression, and a minimally adjusted sensitivity model for work disruption, controlling for age and education.

Results: Female veterans had lower rates of high financial literacy than male veterans (15/29, 52% vs 48/59, 81%; *P*=.006) and lower correct-response rates on compound interest (10/29, 35% vs 36/59, 61%; *P*=.02) and inflation (14/29, 48% vs 43/59, 73%; *P*=.03). Black veterans had lower correct-response rates than non-Black veterans on inflation (11/24, 46% vs 46/64, 72%; *P*=.03) and retirement strategy (15/24, 63% vs 56/64, 88%; *P*=.014), although composite high-literacy rates did not differ significantly by race. In unadjusted models among participants with complete outcome data (n=75), lower financial literacy was directionally associated with higher odds of all 4 financial toxicity outcomes, with the clearest association observed for work disruption (odds ratio 0.56 per 1-point increase in financial literacy score, 95% CI 0.33-0.95; *P*=.03). Black female veterans reported elevated financial toxicity across multiple domains. Financial support program use was low overall (29%).

Conclusions: These findings suggest that financial literacy may be a marker of vulnerability to financial toxicity among veterans, but observed associations should be regarded as preliminary and hypothesis-generating. The results identify concept-level financial literacy domains and work disruption as candidate signals for future screening evaluation. Future research should evaluate whether brief screening, financial literacy assessment, and benefit-navigation strategies improve identification, referral, adherence, and downstream financial and health-related outcomes in larger and more representative veteran populations.

(Online J Public Health Inform 2026;18:e97291) doi:10.2196/97291

KEYWORDS

financial toxicity; veterans; public health informatics; screening; digital health; health equity; social determinants of health; risk stratification; benefit navigation

Introduction

Financial toxicity, defined as the financial strain arising from health care costs and its downstream material, psychological, and behavioral consequences, was initially recognized and operationalized in oncology research. Cancer care provided an early setting for this concept because patients often face prolonged treatment courses, high out-of-pocket expenses, insurance complexity, employment disruption, and difficult trade-offs between treatment adherence and household financial stability [1-3]. Early oncology work framed financial toxicity as a patient-reported outcome because financial strain could affect not only material well-being but also quality of life, symptom burden, treatment decisions, and adherence [2,3].

XSL•FO
RenderX

Although the construct originated in cancer care, similar mechanisms may operate in other populations facing chronic illness, fragmented benefits, employment disruption, and health care affordability barriers.

Among US veterans, financial well-being is closely linked to post-service health, housing stability, and reintegration into civilian life [4-7]. Prior research has documented substantial heterogeneity in financial outcomes within veteran populations, with disparities across race, sex, and socioeconomic status [8,9]. Female veterans face distinct structural and economic challenges, including lower earnings and higher unmet health care needs compared with male veterans [10,11]. Veterans of color are also more likely to experience housing instability and adverse social determinants of health, which may negatively affect both financial and health outcomes [12,13].

Financial literacy, defined as the ability to understand and apply core economic concepts such as compounding, inflation, and investment diversification, plays an important role in financial decision-making and long-term economic resilience [14]. Lower financial literacy has been associated with a greater debt burden, poorer financial behaviors, and heightened vulnerability to financial shocks [15,16]. In health care settings, these vulnerabilities may exacerbate financial toxicity when individuals face income disruption, out-of-pocket medical expenses, or complex benefit systems. However, the relationship between specific financial literacy concept deficits and health-related financial toxicity has not been empirically examined in veteran populations. In this study, financial literacy refers specifically to 3 benchmark financial concepts assessed in the National Financial Capability Study (NFCS)—compound interest, inflation, and retirement diversification—rather than broader financial capability or health-insurance literacy.

Prior work has documented racial and sex-based differences in financial capability among US veterans using large national samples [8,9]. These studies showed that female veterans and veterans of color reported lower financial capability scores than their White male counterparts. However, prior analyses relied largely on composite financial capability measures and did not assess whether lower performance on specific financial literacy concepts was associated with health-related financial toxicity outcomes such as treatment-adherence impact, work disruption, or borrowing for care. As a result, it remains unclear whether financial literacy gaps and financial toxicity burdens reflect similar or different patterns across veteran subgroups. Whether observed subgroup differences in literacy reflect independent racial and gender effects or are partially attributable to differences in educational attainment or age composition within this convenience sample warrants careful interpretation. A concept-level approach may provide more actionable insight by identifying which domains of financial knowledge appear most limited among veterans. Emerging public health frameworks suggest that examining race and sex independently may obscure compounded vulnerabilities experienced by specific subgroups [17,18]. Therefore, exploratory subgroup analysis may help identify patterns that warrant further study in larger and more representative veteran samples.

This study examines concept-level financial literacy and financial toxicity indicators among US veterans and explores preliminary associations between financial literacy and health-related financial strain. We examine subgroup patterns by race and self-reported sex, including an exploratory analysis of the Black female subgroup, descriptively to identify areas for future research rather than to draw confirmatory conclusions. We hypothesize that financial literacy differs across racial and sex subgroups and that lower financial literacy is associated with higher financial toxicity. Beyond describing exploratory subgroup patterns, this study identifies candidate indicators that could inform future evaluations of public health informatics approaches for brief financial toxicity screening, risk stratification, and benefit-navigation support for veterans facing health care affordability barriers.

Methods

Research Objectives

The objectives of this study were to (1) describe financial toxicity indicators among US veterans; (2) evaluate concept-level financial literacy across subgroups defined by race and sex; and (3) explore preliminary associations between financial literacy and financial toxicity, with a minimally adjusted sensitivity analysis for work disruption controlling for age and education.

Study Design and Population

This cross-sectional survey examined financial literacy and financial toxicity among US veterans. Eligible participants were adults aged 18 years or older who self-identified as US veterans. Participants were recruited between July 2024 and April 2025 through an anonymous survey link distributed via veteran-focused community contacts, online veteran networks, and in-person outreach at local community events. No financial incentive was provided. Because recruitment was anonymous and community-based, the number of individuals who received or viewed the survey invitation could not be determined. In-person outreach occurred primarily in the Northern Virginia or Washington, DC, metropolitan region, while online recruitment may have reached veterans outside this area. The survey was administered through an open, anonymous Google Forms link without log-in or collection of direct identifiers; duplicate-entry prevention measures, such as cookies or IP-address checks, were not used.

Individuals who were younger than 18 years and did not identify as veterans were excluded. Additionally, participants with incomplete responses to all 3 financial literacy items were also excluded from the financial literacy analyses. The final analytic sample for financial literacy analyses included 88 participants. Financial toxicity items were completed by 75 of the 88 participants; therefore, analyses involving financial toxicity outcomes were based on this subsample, with valid-response denominators reported as applicable to each outcome.

Participation was voluntary and anonymous. Reporting was guided by the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement [19]. Because this study was exploratory and based on a modest community

XSL•FO
RenderX

sample, no formal a priori power calculation was performed. This was a nonprobability convenience sample and was not designed to be nationally representative of all US veterans. Analyses were intended to be descriptive and hypothesis-generating, with limited power for complex multivariable analyses and small subgroup inference.

Measures

Financial literacy was assessed using 3 benchmark items adapted from the NFCS, covering compound interest, inflation, and retirement strategy. Each item was coded as correct or incorrect, with "don't know" and nonresponse coded as incorrect. Correct responses were defined as: "More than US $1150" for compound interest, "Less" for inflation, and "Investing in a mix of stocks and bonds" for retirement strategy. Scores were summed to create a composite financial literacy score ranging from 0 to 3, with higher scores indicating greater financial literacy. For descriptive and logistic analyses, financial literacy was additionally dichotomized as low financial literacy (0‑1 correct) vs high financial literacy (2‑3 correct). These 3 items were selected because they draw on the core concepts assessed by the widely used "Big Three" financial literacy questions [14]—interest compounding, inflation, and risk diversification (here operationalized as retirement or investment diversification). The full NFCS instrument was not administered to reduce respondent burden in a brief community-based veteran survey and to focus on the concepts plausibly relevant to financial vulnerability and health-related financial strain. The resulting score should therefore be interpreted as a brief concept-level literacy measure rather than a comprehensive assessment of financial capability.

Financial toxicity was assessed using items adapted from and aligned with the domains reflected in the Comprehensive Score for Financial Toxicity (COST) framework, rather than the validated COST instrument itself, including difficulty affording care, reduced or quit work due to health, borrowing money or using savings for care, and treatment-adherence impact. The full COST instrument was not used because it was developed and validated for patients undergoing active cancer treatment, whereas this study involved a brief community-based survey of a nononcology veteran population. The 4 items were chosen to capture practical, patient-reported manifestations of health-related financial strain that may be relevant to future screening workflows—difficulty affording care, work disruption, borrowing or using savings, and treatment-adherence impact—while limiting respondent burden. These adapted items should not be interpreted as producing a validated COST score. Program use was assessed separately as reported use of available financial support resources. Thirteen participants did not answer the financial toxicity items; therefore, financial toxicity outcomes were available for 75 of 88 participants. The exact wording of all financial literacy and financial toxicity survey items, including response options and coding rules, is provided in Multimedia Appendix 1. Reporting details for the online survey component are provided in the completed CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist (Checklist 1).

Demographic variables included age category, race and ethnicity, sex, education level, and marital status. Race and ethnicity was collected as a single self-reported category, with Hispanic analyzed as its own category. Sex was self-reported using male and female response options; gender identity was not separately assessed. Because the survey captured only binary self-reported sex, results involving this variable are reported in terms of sex, and gender identity–based interpretations are beyond the scope of these data. Race and sex were examined because prior veteran literature has documented disparities in financial capability and financial vulnerability across these groups. For subgroup analyses, race comparisons focused primarily on Black vs non-Black participants because of sample size constraints. The non-Black comparison group included participants who identified as White, Hispanic, Asian or Pacific Islander, other race and ethnicity, or preferred not to answer. This grouping was used only for exploratory analysis and should not be interpreted as implying that non-Black participants are a homogeneous group. Intersectional analyses were interpreted cautiously because subgroup counts were small. Health care setting and coverage variables, including Veterans Health Administration use, Medicare, Medicaid, private insurance, other insurance status, and payer source, were not assessed.

Statistical Analysis

Descriptive statistics were used to summarize participant characteristics, financial literacy, and financial toxicity outcomes. Percentages were calculated using valid-response denominators. Group differences in categorical outcomes were evaluated using Fisher exact tests since cell sizes were small. All hypothesis tests were 2-sided. Given the exploratory nature of the analyses and the number of comparisons performed, *P* values are reported descriptively without correction for multiple testing; findings should not be interpreted as confirmatory.

Financial literacy was examined in two ways: (1) as a composite high-vs-low literacy variable and (2) at the item level for compound interest, inflation, and retirement strategy. Composite and item-level literacy differences were compared across sex and race subgroups. Due to small subgroup counts, Black female veterans were examined descriptively and inferentially only in limited exploratory comparisons.

Associations between financial literacy and each financial toxicity outcome were examined using unadjusted logistic regression in 2 specifications: binary financial literacy status (low vs high) and continuous financial literacy score (0‑3). Separate unadjusted logistic regression models were run for each financial toxicity outcome: treatment-adherence impact, reduced or quit work, borrowing money or using savings for care, and difficulty affording care. Odds ratios (ORs) in continuous models were interpreted per 1-point increase in financial literacy score. Because work disruption emerged as the clearest unadjusted signal, and because age and education were considered plausible confounders of both financial literacy and financial strain, we additionally conducted a minimally adjusted sensitivity analysis for the reduced-or-quit-work outcome controlling for age category and education level. Given the modest sample size and limited number of work-disruption events, age and education were modeled parsimoniously as

XSL•FO
RenderX

ordered categories to reduce overfitting. All regression findings were interpreted as exploratory and hypothesis-generating.

Statistical code used for the analyses is available from the corresponding author on reasonable request.

Ethical Considerations

The study was reviewed by the WCG Institutional Review Board and determined to be exempt (IRB 2155779) on November 27, 2024. The study was conducted in accordance with applicable ethical standards for human subjects research. Informed consent was obtained from all participants involved in the study through voluntary anonymous survey participation after the presentation of study information. Consistent with the exemption determination, formal written informed consent was not required; however, participants were informed of the study's purpose, voluntary nature, and anonymity of their responses prior to completing the survey. No individual person's data, images, or videos are included in this manuscript.

Results

Participant Characteristics

The final analytic sample included 88 US veterans, as summarized in Table 1. Of the 88 participants, the sample comprised 59 (67%) men and 29 (33%) women. By race and ethnicity, 51 (58%) participants identified as White, 24 (27%) as Black, 7 (8%) as Hispanic, and 4 (5%) as Asian or Pacific Islander; 2 (2%) participants identified as another category or preferred not to answer. Age was reported in categories, with the largest proportions in the 30 to 40 and 50 to 60 year groups. Most participants reported at least some college education (80/88, 91%). The majority were married or partnered (60/88, 68%), while 16 of 88 (18%) were single, and 12 of 88 (14%) were divorced or separated.

XSL•FO
RenderX

Table . Characteristics of included participants (N=88)[a].

Characteristics	Participants, n (%)
Age (y)	
18 - 30	7 (8)
30 - 40	20 (23)
40 - 50	16 (18)
50 - 60	20 (23)
60 - 70	14 (16)
≥70	11 (13)
Self-reported sex	
Male	59 (67)
Female	29 (33)
Race and ethnicity	
White	51 (58)
Black	24 (27)
Hispanic	7 (8)
Asian or Pacific Islander	4 (5)
Other	2 (2)
Education	
Graduate degree (masters or higher)	39 (44)
College or undergraduate	24 (27)
Some college	17 (19)
High school	8 (9)
Marital status	
Married or partnered	60 (68)
Single	16 (18)
Divorced or separated	12 (14)
Race-by-sex subgroups	
White	
Male	34 (39)
Female	17 (19)
Black	
Male	18 (21)
Female	6 (7)
Hispanic	
Male	4 (5)
Female	3 (3)
Asian or Pacific Islander	
Male	2 (2)
Female	2 (2)
Other	
Male	1 (1)
Female	1 (1)

XSL•FO
RenderX

[a]Percentages may not sum to 100% due to rounding.

Financial Literacy by Race and Self-Reported Sex

Composite and item-level financial literacy findings are shown in Tables 2 and 3. In this exploratory sample, female veterans showed lower observed rates of high financial literacy than male veterans (15/29, 52% vs 48/59, 81%; Fisher exact test *P*=.006). At the item level, women also had lower correct-response rates for compound interest (10/29, 35% vs 36/59, 61%; *P*=.02) and inflation (14/29, 48% vs 43/59, 73%; *P*=.03), while retirement strategy did not differ at conventional thresholds by self-reported sex. All sex-based comparisons are similarly reported descriptively, given the exploratory design and the absence of correction for multiple testing.

Table . Composite high financial literacy by demographic subgroup[a].

Comparison	Group 1, n/N (%)	Group 2, n/N (%)	Fisher exact test *P* value
Male vs female	Male: 48/59 (81)	Female: 15/29 (52)	.006
Black vs non-Black	Black: 15/24 (63)	Non-Black: 48/64 (75)	.29
Black women vs all others	Black female: 2/6 (33)	All others: 61/82 (74)	.05

[a]Fisher exact tests were used throughout. High financial literacy was defined as 2 to 3 correct responses out of 3 benchmark financial literacy items.

Table . Financial literacy by item-level correct responses.

Item	Black, n/N (%)	Non-Black, n/N (%)	Fisher exact test *P* value	Female, n/N (%)	Male, n/N (%)	Fisher exact test *P* value
Compound interest	10/24 (42)	36/64 (56)	.24	10/29 (34)	36/59 (61)	.02
Inflation	11/24 (46)	46/64 (72)	.03	14/29 (48)	43/59 (73)	.03
Retirement strategy	15/24 (63)	56/64 (88)	.01	21/29 (72)	50/59 (85)	.25

Black veterans had lower composite high-literacy rates than non-Black veterans (15/24, 63% vs 48/64, 75%), although this difference did not reach conventional thresholds (*P*=.29). At the item level, Black veterans had lower correct-response rates on inflation (11/24, 46% vs 46/64, 72%; *P*=.03) and retirement strategy (15/24, 63% vs 56/64, 88%; *P*=.01), while the difference in compound interest did not reach conventional thresholds (*P*=.24). All item-level comparisons are reported descriptively, given the exploratory design and absence of correction for multiple testing.

Among Black female veterans, 2 of 6 (33%) met the high-literacy threshold, compared with 61 of 82 (74%) among all other participants (Fisher exact test *P*=.052). This comparison is exploratory and descriptive only, given the small subgroup size. These findings are summarized in Tables 2 and 3.

Financial Toxicity Indicators

Financial toxicity outcomes are summarized in Table 4 using valid-response denominators among the 75 participants who completed the financial toxicity items. Overall, 12 of 75 (16%) participants reported difficulty affording treatment or medications, 19 of 75 (25%) reported reduced or quit work because of health, 28 of 75 (37%) reported treatment-adherence impact, and 14 of 75 (19%) reported borrowing money or using savings for care. Financial support program use was reported by 22 of 75 (29%) participants.

Table . Financial toxicity indicators: overall sample (N=75 with complete data).

Indicator	n (%)[a]
Medical expenses <10% of income	52 (69)
Treatment adherence impact (any)	28 (37)
Reduced or quit work	19 (25)
Difficulty affording care	12 (16)
Borrowed or used savings for care	14 (19)
Financial support program use	22 (29)

[a]Denominators reflect valid responses among the 75 participants with complete financial toxicity data.

The work-disruption indicator may not have been applicable to all respondents, for example, those who were unemployed or already retired at the time of the survey, and the denominator therefore includes potentially nonapplicable participants which may affect prevalence estimates and should be considered when interpreting the work-disruption association.

XSL•FO
RenderX

Association Between Financial Literacy and Financial Toxicity

Cross-tabulations and unadjusted logistic regression results are presented inTables 5-7. Veterans with low financial literacy (0 - 1 correct) had numerically higher rates of all 4 financial toxicity outcomes than those with high financial literacy (2 - 3 correct), although not all comparisons reached conventional thresholds. Treatment-adherence impact was reported by 9 of 22 (41%) participants with low financial literacy compared with 19 of 53 (36%) participants with high financial literacy. Reduced or quit work was reported by 9 of 22 (41%) participants with low financial literacy compared with 10 of 53 (19%) participants with high financial literacy. Borrowing money or using savings for care was reported by 6 of 22 (27%) participants with low financial literacy compared with 8 of 53 (15%) participants with high financial literacy. Difficulty affording care was reported by 5 of 22 (23%) participants with low financial literacy compared with 7 of 53 (13%) participants with high financial literacy.

Table . Financial toxicity indicators by financial literacy levels[a].

Outcome	Low financial literacy, 0 - 1 correct, n/N (%)	High financial literacy, 2 - 3 correct, n/N (%)
Treatment-adherence impact	9/22 (41)	19/53 (36)
Reduced or quit work	9/22 (41)	10/53 (19)
Borrowed money or used savings for care	6/22 (27)	8/53 (15)
Difficulty affording care	5/22 (23)	7/53 (13)

[a]Low financial literacy was defined as 0-1 correct responses, and high financial literacy was defined as 2-3 correct responses.

Table . Financial toxicity indicators by unadjusted logistic regression results.

Outcome[a]	OR[b] for low vs high financial literacy (95% CI; *P* value)	OR per 1-point increase in financial literacy score (95% CI; *P* value)
Treatment-adherence impact	1.24 (0.45 - 3.43; *P*=.68)	0.82 (0.51 - 1.32; *P*=.42)
Reduced or quit work	2.98 (1.00 - 8.88; *P*=.051)	0.56 (0.33 - 0.95; *P*=.03)
Borrowed money or used savings for care	2.11 (0.63 - 7.02; *P*=.22)	0.59 (0.33 - 1.05; *P*=.07)
Difficulty affording care	1.93 (0.54 - 6.92; *P*=.31)	0.73 (0.40 - 1.34; *P*=.32)

[a]Unadjusted logistic regression models were run separately for each financial toxicity outcome.

[b]OR: odds ratio.

Table . Financial toxicity indicators by minimally adjusted work-disruption sensitivity model[a].

Predictor	Adjusted OR[b] (95% CI)	*P* value
Financial literacy score, per 1-point increase	0.55 (0.28 - 1.08)	.08
Age category	1.75 (1.09 - 2.82)	.02
Education level	0.41 (0.22 - 0.77)	.005

[a]Data show minimally adjusted sensitivity model for reduced or quit work, controlling for age category and education level. Age and education were modeled as ordered categories. N=75 reflects participants who completed the financial toxicity items.

[b]OR: odds ratio.

In unadjusted logistic models using low vs high financial literacy, lower financial literacy was directionally associated with higher odds of each financial toxicity outcome. The strongest unadjusted signal was observed for work disruption (OR 2.98, 95% CI 1.00 - 8.88; *P*=.051), which did not reach conventional thresholds in the binary model. In continuous score models, each 1-point increase in financial literacy score was associated with lower odds of reduced or quit work (OR 0.56, 95% CI 0.33 - 0.95; *P*=.03), though this finding should be interpreted descriptively given the exploratory design and absence of multiple testing correction. Associations for treatment-adherence impact, borrowing or savings use, and difficulty affording care were directionally similar but did not reach conventional thresholds. In a minimally adjusted sensitivity model controlling for age category and education level, the association between continuous financial literacy score and reduced or quit work remained directionally similar but was attenuated and did not reach conventional thresholds (adjusted OR 0.55, 95% CI 0.28 - 1.08; *P*=.08). Within the same minimally adjusted model, higher education level was independently associated with lower odds of reduced or quit work (adjusted OR, 0.41; 95% CI, 0.22 - 0.77; *P*=.005), whereas older age category was associated with higher odds of reduced or quit work (adjusted OR 1.75, 95% CI 1.09 - 2.82; *P*=.02). The corresponding minimally adjusted model using dichotomous financial literacy status showed a similar pattern (adjusted OR for low vs high financial literacy, 2.90, 95% CI 0.71 - 11.75; *P*=.13).

XSL•FO
RenderX

As a secondary sensitivity analysis, we examined an alternative high-literacy threshold consistent with prior NFCS-based research. Because high financial literacy is often defined as 3 of 3 correct responses in NFCS-based work, we conducted a sensitivity analysis using that stricter threshold. Results for work disruption were directionally similar, with higher odds among participants scoring 0 to 2 correct compared with those scoring 3 of 3 correct, but the association did not reach conventional thresholds (OR 2.26, 95% CI 0.72 - 7.12; *P*=.17). In a minimally adjusted model controlling for age category and education level, the direction was unchanged but remained imprecise (adjusted OR 2.53, 95% CI 0.69 - 9.32; *P*=.16).

Subgroup Patterns in Financial Toxicity

Exploratory subgroup comparisons are presented in Table 8. In descriptive analyses, Black female veterans with available data had numerically higher reported financial toxicity across all 4 assessed financial toxicity outcomes. Among Black female veterans with available data, 3 of 5 (60%) reported difficulty affording care, 4 of 5 (80%) reported reduced or quit work, 3 of 5 (60%) reported borrowing money or using savings for care, and 5 of 5 (100%) reported treatment-adherence impact.

Table . Financial toxicity indicators by race and sex subgroup (valid-response denominators)[a].

Subgroup	N (toxicity data)	Difficulty affording care, n/N (%)	Reduced work, n/N (%)	Borrowed or savings, n/N (%)	Adherence impact, n/N (%)	Program use, n/N (%)
White male	33	2/33 (6)	5/33 (15)	3/33 (9)	6/33 (18)	11/33 (33)
White female	17	3/17 (18)	4/17 (24)	4/17 (24)	9/17 (53)	5/17 (29)
Black male	7	3/7 (43)	5/7 (71)	1/7 (14)	4/7 (57)	3/7 (43)
Black female	5	3/5 (60)	4/5 (80)	3/5 (60)	5/5 (100)	0/5 (0)
Hispanic male	4	1/4 (25)	1/4 (25)	2/4 (50)	3/4 (75)	1/4 (25)
Hispanic female	3	0/3 (0)	0/3 (0)	0/3 (0)	1/3 (33)	0/3 (0)
Asian male	2	0/2 (0)	0/2 (0)	0/2 (0)	0/2 (0)	0/2 (0)
Asian female	2	0/2 (0)	0/2 (0)	0/2 (0)	0/2 (0)	0/2 (0)
Other female	1	0/1 (0)	1/1 (100)	1/1 (100)	0/1 (0)	0/1 (0)

[a]Percentages are based on item-specific valid-response denominators. All subgroup counts were small; all results are descriptive only and should not be interpreted as confirmatory. For reference, Fisher exact comparisons of Black female veterans vs all other participants yielded *P*=.04 for difficulty affording care, *P*=.01 for reduced work, *P*=.04 for borrowed money or use of savings, and *P*=.006 for treatment-adherence impact; given n=5, these values are not reliable for inference. Program use did not differ at conventional thresholds (*P*=.31).

In a descriptive comparison limited to female veterans with available financial toxicity data, Black female veterans also had numerically higher rates of all 4 assessed financial toxicity outcomes than other female veterans. Difficulty affording care was reported by 3 of 5 (60%) Black female veterans compared with 3 of 23 (13%) other female veterans, reduced or quit work by 4 of 5 (80%) vs 5 of 23 (22%), borrowing money or using savings for care by 3 of 5 (60%) vs 5 of 23 (22%), and treatment-adherence impact by 5 of 5 (100%) vs 10 of 23 (43%). Financial support program use was reported by 0 of 5 (0%) Black female veterans compared with 5 of 23 (22%) other female veterans. These comparisons are descriptive only due to the very small number of Black female veterans with complete financial toxicity data.

Given the very small cell size (n=5), these comparisons are descriptive only, and no inferential conclusions should be drawn. Fisher exact test *P* values are reported in the table footnote for completeness but should not be interpreted as statistically reliable. Financial support program use was zero among Black female veterans with available data.

Among White veterans, women reported numerically higher rates than men for 4 of 5 financial toxicity indicators, although subgroup counts were small, and these comparisons were descriptive only. Financial support program use was lower among White female veterans than among White male veterans.

Discussion

Principal Findings

In this exploratory survey of 88 US veterans, 4 principal findings emerged. First, female veterans had lower financial literacy than male veterans, most notably regarding compound interest and inflation. Second, Black veterans showed lower correct-response rates than non-Black veterans for inflation and retirement strategy, although composite high-literacy rates did not differ significantly by race. Third, lower financial literacy was directionally associated with higher financial toxicity across all 4 assessed outcomes in unadjusted analyses, with the clearest signal observed for work disruption; however, the work-disruption association was attenuated and no longer reached conventional thresholds after minimal adjustment for age and education. Fourth, financial literacy gaps and financial toxicity burden appeared to follow partially distinct demographic patterns, with the numerically highest reported toxicity burden observed among Black female veterans in this small exploratory subgroup.

Female veterans demonstrated lower financial literacy than their male counterparts on both composite and item-level measures, with differences most pronounced in compound interest and inflation knowledge. These findings are consistent with prior work documenting sex-based differences in financial capability

XSL•FO
RenderX

among veterans [8,9] and with broader literature on sex-based differences in financial literacy [14-16]. This study builds on prior work by suggesting specific concept domains that contributed to the observed differences rather than examining financial literacy as a single aggregate score. Compound interest and inflation knowledge may be especially relevant to health-related financial decision-making, as inflation affects the real value of fixed incomes and veterans' benefits over time, while compound interest affects the long-term cost of borrowing for medical care. Retirement strategy knowledge may similarly affect long-term financial resilience and the capacity to absorb health care–related income shocks without depleting savings.

Black veterans showed lower correct-response rates than non-Black veterans specifically for inflation and retirement strategy, although overall composite high-literacy rates did not differ at conventional thresholds by race. It should be noted that observed item-level differences between Black and non-Black veterans were not adjusted for educational attainment or age composition, and it remains possible that these factors partially account for the observed patterns. The item-level approach offers more granular insight than composite measures and may help identify potential targets for future education or navigation efforts that may not be apparent from aggregate scores alone.

In unadjusted logistic models, lower financial literacy was directionally associated with higher odds of all 4 assessed financial toxicity outcomes. The clearest unadjusted association was observed for work disruption, where each 1-point increase in financial literacy score was associated with lower odds of reducing or quitting work. However, in a minimally adjusted sensitivity model controlling for age category and education level, this association was attenuated and no longer reached conventional thresholds. The other 3 outcomes showed similar directional patterns in unadjusted analyses but did not reach conventional thresholds. Taken together, these findings suggest a possible relationship between lower financial literacy and health-related financial strain, but they do not establish an independent association in this sample. The observed attenuation after adjustment and the small sample size underscore the need for larger, adequately powered studies with multivariable modeling. Furthermore, the sample's high educational attainment (91% with at least some college) is not representative of the broader veteran population and may have influenced estimates of financial strain, meaning prevalence estimates in broader veteran populations may differ from those observed here. Separately, financial toxicity indicators were observed even though many participants reported medical expenses below 10% of income. This pattern is consistent with the possibility that health-related financial strain may reflect not only direct out-of-pocket spending but also broader pressures related to employment disruption, savings depletion, and care access.

Finally, financial literacy gaps and financial toxicity burdens appeared to follow partially different demographic patterns in this sample. For financial literacy, the clearest composite difference was observed by sex, with female veterans demonstrating lower high-literacy rates than male veterans. In contrast, the highest financial toxicity burden was observed among Black female veterans, who comprised a small exploratory subgroup. Among Black female veterans with available data, elevated rates were observed across all 4 assessed financial toxicity outcomes. This pattern should not be interpreted as evidence that all female veterans had uniformly higher financial toxicity than male veterans. Rather, it suggests that financial literacy gaps and financial toxicity may reflect overlapping but distinct sources of vulnerability. Financial literacy interventions targeting female veterans may address some knowledge gaps but may not fully address the structural, economic, health care access, and benefit-navigation factors that could contribute to financial toxicity among Black female veterans. These findings are preliminary and require confirmation in adequately powered intersectional analyses.

Comparison With Prior Work

These findings are broadly consistent with, but extend, prior national-level work by Mottola and Skimmyhorn [8] and Skimmyhorn et al [9], which documented financial capability disparities by race and sex among veterans. Furthermore, this study adds exploratory evidence in 3 areas not captured in those reports: concept-level assessment of specific financial literacy gaps, preliminary examination of the association between financial literacy and health-related financial toxicity, and descriptive documentation of financial toxicity patterns, including treatment-adherence impact and work disruption, in a veteran population outside of a clinical disease context.

Beyond the veteran context, these findings highlight a broader public health concern: financial strain related to health care does not arise solely from medical need but also from the interaction of economic resources, financial knowledge, and access to supportive systems. In that sense, the present results may have relevance for other vulnerable populations facing cost-related barriers to care, employment disruption, or difficulty navigating complex benefit structures [20-22]. The concept-level approach used here may therefore be useful not only for veteran populations but also for identifying actionable financial literacy gaps in other groups at risk of health-related financial strain [14-16].

From a public health and policy perspective, these findings suggest potential value in integrating brief financial toxicity screening into veteran care workflows, particularly for indicators such as work disruption and treatment-adherence strain. Targeted financial literacy education addressing core concepts, particularly compound interest, inflation, and retirement strategy, may be more focused than generalized financial education and may warrant further evaluation, particularly among female veterans. Such efforts may be strengthened with active benefit-navigation support, as financial support program utilization was low (29%) and did not differ significantly even among the subgroup with the highest reported toxicity burden.

Public Health Informatics Implications

These findings may inform future evaluations of brief informatics-enabled financial toxicity screening workflows for veterans, with particular attention to work disruption, borrowing money or using savings for care, difficulty affording care, treatment-adherence impact, and low financial support program utilization. However, implementation would likely differ across care environments. In the Veterans Health Administration or

XSL•FO
RenderX

other integrated systems, these domains could potentially be incorporated into electronic health record–based screening tools, patient portal questionnaires, or referral workflows. In community-based, private, Medicare, Medicaid, or uninsured or underinsured care settings, similar screening may require lower-resource approaches, such as brief paper or digital questionnaires, community organization referral pathways, social work referral protocols, or partnerships with veteran service organizations. Because this study did not assess Veterans Health Administration use, insurance type, payer source, or care coordination setting, we could not evaluate how the health care environment influenced financial toxicity or the feasibility of screening implementation. Future research should compare screening, referral, and benefit-navigation approaches across Veterans Health Administration and community-based care settings and assess whether these workflows improve identification, referral completion, treatment adherence, and downstream financial and health outcomes.

Strengths and Limitations

This study has several notable strengths. It uses standardized NFCS benchmark items to assess concept-level financial literacy, enabling comparison with prior national-level research. By analyzing financial literacy at the concept level rather than only as a composite score, the study suggests specific knowledge domains—particularly compound interest, inflation, and retirement strategy—that may warrant further evaluation in future education or navigation interventions. The simultaneous examination of financial literacy and health-related financial toxicity within the same veteran sample addresses an important gap in the existing literature, which has generally examined these constructs separately. In addition, applying a COST-aligned framework to a nononcology veteran population extends this line of inquiry beyond its usual clinical context and captures patient-reported outcomes, including treatment-adherence impact and work disruption, that have not been extensively assessed in prior veteran financial capability research.

Several limitations should be acknowledged. First, the modest sample size (N=88; n=75 for financial toxicity analyses) limited statistical power, especially for subgroup analyses and multivariable modeling. Race-by-sex subgroups were particularly small, including Black female veterans (n=6) with complete financial toxicity data (n=5), and those findings should therefore be interpreted cautiously. Second, the study used a convenience sample that was also highly educated, with 91% (n=80) of participants reporting at least some college and 44% (n=39) reporting a graduate degree. This educational profile is not representative of the broader US veteran population, limits external validity, and may have influenced estimates of financial strain in this sample. Participant ZIP code, Area Deprivation Index, Social Vulnerability Index, Veterans Health Administration use, insurance type, payer source, and care coordination setting were not collected; therefore, we could not evaluate whether neighborhood socioeconomic context, health care coverage setting, or care environment influenced financial toxicity, financial support use, or the feasibility of screening and navigation workflows. Third, although a minimally adjusted sensitivity model was conducted for work disruption, residual confounding by income, marital status, employment context, and other unmeasured factors remains possible. Fourth, the cross-sectional design precludes causal inference and leaves open the possibility of reverse causation. Fifth, financial toxicity was assessed using a framework-aligned adaptation rather than the fully validated COST instrument. Sixth, findings involving dichotomized financial literacy were sensitive to threshold choice; when high financial literacy was defined more strictly as 3 of 3 correct responses, the direction of association with work disruption was similar, but estimates were attenuated and remained imprecise. The pragmatic 0‑1 vs 2‑3 threshold used in primary analyses is not the most common cut point in NFCS literature, and results should be interpreted with this in mind.

Several individual-level factors known to shape financial literacy and financial decision-making in aging populations were not assessed. We did not measure cognitive performance, frailty, neurobiological correlates, functional status, or financial exploitation vulnerability, each of which may be relevant to financial literacy and financial vulnerability in older adults. Lower performance on benchmark financial literacy concepts has been associated with cognitive impairment and structural brain differences in older adults [23], and frailty has been proposed as a correlate of perceived financial exploitation in older women [24]. Because a substantial proportion of this sample was older, these unmeasured factors may have contributed to the observed financial literacy and financial toxicity patterns, and we did not examine whether sex-based differences in financial literacy varied by age. Relatedly, heightened financial vulnerability documented among older Black women in other work [25] reinforces the value of examining intersectional subgroups with richer individual-level data than were available here. Finally, the 3 financial literacy items capture widely used benchmark concepts but do not cover all domains that may influence financial toxicity, such as debt literacy, insurance literacy, budgeting, credit use, benefit navigation, and health care cost understanding. Because participants were recruited within a US veteran context, findings should be generalized cautiously to non-US populations and health care systems with different financing structures.

Conclusions

This exploratory study observed subgroup differences in specific financial literacy concepts among US veterans. Female veterans showed lower composite and item-level financial literacy than male veterans, with lower correct-response rates most evident for compound interest and inflation. Black veterans showed lower correct-response rates on inflation and retirement strategy than non-Black veterans, although composite high-literacy rates did not differ significantly by race. In unadjusted analyses, lower financial literacy was directionally associated with higher reported financial toxicity across all 4 assessed outcomes, with the clearest signal observed for work disruption; however, this association was attenuated after minimal adjustment for age and education. Black female veterans reported numerically elevated financial toxicity across all 4 assessed outcomes in a very small subgroup, and these findings should be interpreted cautiously.

XSL•FO
RenderX

Beyond describing exploratory subgroup patterns, this study highlights specific candidate data elements that could support future public health informatics approaches to detecting financial toxicity in veterans and other populations facing health care affordability challenges. Concept-level financial literacy gaps, work disruption, treatment-adherence impact, and low use of support programs may warrant evaluation as brief digital screening indicators and subgroup-aware risk-stratification variables. Future studies should evaluate whether integrating these indicators into electronic health record–based screening workflows, patient portal questionnaires, referral dashboards, and community-facing paper or digital referral tools could help clinicians and health systems identify financially vulnerable patients earlier and connect them to appropriate navigation and support services. Future research should also test whether such informatics-enabled strategies improve identification, referral completion, treatment adherence, and downstream financial and health-related outcomes in larger veteran cohorts and, where relevant, in other high-risk groups.

Acknowledgments

The authors thank the veterans who participated in this study and the community organizations that assisted with outreach and recruitment. The authors used ChatGPT (OpenAI) for language editing and manuscript revision. All final content was reviewed and approved by the authors.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated and analyzed during this study are openly available in Figshare [26]. The statistical code used for the analyses is available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: HS, RR
Data curation: HS
Formal analysis: HS
Investigation: HS
Methodology: HS, RR
Supervision: RR
Writing – original draft preparation: HS
Writing – review and editing: HS, RR

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey items, response options, and coding rules used to assess financial literacy and financial toxicity.
[DOCX File, 20 KB - ojphi_v18i1e97291_app1.docx]

Checklist 1
CHERRIES checklist.
[DOCX File, 22 KB - ojphi_v18i1e97291_app2.docx]

References

1. Financial toxicity (financial distress) and cancer treatment (PDQ®)–patient version. National Cancer Institute. URL: https://www.cancer.gov/about-cancer/managing-care/track-care-costs/financial-toxicity-pdq [accessed 2025-10-06]
2. de Souza JA, Yap BJ, Hlubocky FJ, et al. The development of a financial toxicity patient-reported outcome in cancer: the COST measure. Cancer 2014 Oct 15;120(20):3245-3253. [doi: 10.1002/cncr.28814] [Medline: 24954526]
3. de Souza JA, Yap BJ, Wroblewski K, et al. Measuring financial toxicity as a clinically relevant patient-reported outcome: the validation of the COmprehensive Score for financial Toxicity (COST). Cancer 2017 Feb 1;123(3):476-484. [doi: 10.1002/cncr.30369] [Medline: 27716900]
4. Hoffmire CA, Borowski S, Vogt D. Contribution of veterans' initial post-separation vocational, financial, and social experiences to their suicidal ideation trajectories following military service. Suicide Life Threat Behav 2023 Jun;53(3):443-456. [doi: 10.1111/sltb.12955] [Medline: 36919380]

XSL•FO
RenderX

5. Elbogen EB, Johnson SC, Wagner HR, Newton VM, Beckham JC. Financial well-being and postdeployment adjustment among Iraq and Afghanistan war veterans. Mil Med 2012 Jun;177(6):669-675. [doi: 10.7205/milmed-d-11-00388] [Medline: 22730842]
6. Demographic shifts and higher housing costs are leading to more cost burdens among veteran renters. National Low Income Housing Coalition. URL: https://nlihc.org/resource/demographic-shifts-and-higher-housing-costs-are-leading-more-cost-burdens-among-veteran [accessed 2025-10-20]
7. Elbogen EB, Zeber JE, Vogt D, Perkins DF, Finley EP, Copeland LA. Financial status and well-being in recently separated military veterans. Mil Med 2023 Jul 22;188(7-8):usac030. [doi: 10.1093/milmed/usac030] [Medline: 35253060]
8. Mottola G, Skimmyhorn W. How are veterans faring financially? : FINRA Investor Education Foundation; 2019 URL: https://www.finrafoundation.org/sites/finrafoundation/files/how-are-veterans-faring-financially-issue-brief_0.pdf [accessed 2025-10-20]
9. Skimmyhorn W, Mottola GR, Valdes O. The financial capability of United States military veterans: the role of gender and racial/ethnic identity. : FINRA Investor Education Foundation; 2023 URL: https://www.finrafoundation.org/sites/finrafoundation/files/2024-10/research-brief-veterans-financial-capability-11-23.pdf [accessed 2026-07-10]
10. Washington DL, Bean-Mayberry B, Riopelle D, Yano EM. Access to care for women veterans: delayed healthcare and unmet need. J Gen Intern Med 2011 Nov;26 Suppl 2(Suppl 2):655-661. [doi: 10.1007/s11606-011-1772-z] [Medline: 21989618]
11. Frayne SM, Phibbs CS, Saechao F, et al. Sourcebook: women veterans in the Veterans Health Administration. Volume 4: longitudinal trends in sociodemographics, utilization, health profile, and geographic distribution. : Women's Health Services, Veterans Health Administration, U.S. Department of Veterans Affairs; 2018 URL: https://www.womenshealth.va.gov/WOMENSHEALTH/docs/Sourcebook_Vol_4.pdf [accessed 2025-07-07]
12. Tsai J, Rosenheck RA. Risk factors for homelessness among US veterans. Epidemiol Rev 2015;37(1):177-195. [doi: 10.1093/epirev/mxu004] [Medline: 25595171]
13. Montgomery AE, Tsai J, Blosnich JR. Demographic correlates of veterans' adverse social determinants of health. Am J Prev Med 2020 Dec;59(6):828-836. [doi: 10.1016/j.amepre.2020.05.024] [Medline: 33220754]
14. Lusardi A, Mitchell OS. The economic importance of financial literacy: theory and evidence. J Econ Lit 2014 Mar;52(1):5-44. [doi: 10.1257/jel.52.1.5] [Medline: 28579637]
15. Lusardi A, Tufano P. Debt literacy, financial experiences, and overindebtedness. J Pension Econ Finance 2015 Oct;14(4):332-368. [doi: 10.1017/S1474747215000232]
16. Fernandes D, Lynch Jr JG, Netemeyer RG. Financial literacy, financial education, and downstream financial behaviors. Manage Sci 2014 Aug;60(8):1861-1883. [doi: 10.1287/mnsc.2013.1849]
17. Bauer GR. Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. Soc Sci Med 2014 Jun;110:10-17. [doi: 10.1016/j.socscimed.2014.03.022] [Medline: 24704889]
18. Bowleg L. The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health. Am J Public Health 2012 Jul;102(7):1267-1273. [doi: 10.2105/AJPH.2012.300750] [Medline: 22594719]
19. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296. [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
20. Tucker-Seeley RD, Li Y, Subramanian SV, Sorensen G. Financial hardship and mortality among older adults using the 1996-2004 Health and Retirement Study. Ann Epidemiol 2009 Dec;19(12):850-857. [doi: 10.1016/j.annepidem.2009.08.003] [Medline: 19944348]
21. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2(Suppl 2):19-31. [doi: 10.1177/00333549141291S206] [Medline: 24385661]
22. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Aff (Millwood) 2002;21(2):60-76. [doi: 10.1377/hlthaff.21.2.60] [Medline: 11900187]
23. Giannouli V. Cognitive and brain volume correlates of financial literacy assessment in Greek persons with amnestic Mild Cognitive Impairment: a preliminary study. Front Dement 2026;5:1730124. [doi: 10.3389/frdem.2026.1730124] [Medline: 42136946]
24. Giannouli V. Letter to the editor: does higher prevalence of frailty in Greek older community-dwelling women also relate to higher prevalence of perceived financial exploitation? A new question to ponder upon. J Frailty Aging 2022;11(4):436-437. [doi: 10.14283/jfa.2022.57] [Medline: 36346732]
25. Giannouli V, Hall L, Lichtenberg PA. Do older black women who were financially exploited share specific psychological, cognitive, social and physical health characteristics? The role of financial exploitation vulnerability. J Appl Gerontol 2026 Apr;45(4):712-719. [doi: 10.1177/07334648251347998] [Medline: 40498995]
26. Financial toxicity as a public health issue: intersectional disparities and underutilized financial support programs among urban U.S. veterans. Figshare. URL: https://doi.org/10.6084/m9.figshare.30535838 [accessed 2026-07-07]

XSL•FO
RenderX

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COST: Comprehensive Score for Financial Toxicity
NFCS: National Financial Capability Study
OR: odds ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by E Mensah; submitted 05.Apr.2026; peer-reviewed by SM Leder, V Giannouli; revised version received 24.Jun.2026; accepted 29.Jun.2026; published 07.Aug.2026.

Please cite as:
Sabharwal H, Raya R
Financial Literacy and Financial Toxicity Among US Veterans: Cross-Sectional Survey Informing Public Health Informatics Screening
Online J Public Health Inform 2026;18:e97291
URL: https://ojphi.jmir.org/2026/1/e97291
doi:10.2196/97291

© Harman Sabharwal, Ramona Raya. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 7.Aug.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Burden of Stroke in China From 1990 to 2023: Analysis From the Global Burden of Disease Study 2023

Lijuan Fu[1], MD; Wencai Jiang[1], MD; Yanhua Peng[1], MD; Xianjie Zhang[1], MD; Rui Zhou[2], MD

[1]Deyang People's Hospital, Deyang, China

[2]Shanghai Fourth People's Hospital, No. 1279, Sanmen Road, Shanghai, China

Corresponding Author:
Rui Zhou, MD
Shanghai Fourth People's Hospital, No. 1279, Sanmen Road, Shanghai, China

Abstract

Background: China has initiated a national stroke program since 2011. Understanding the magnitude of stroke burden in China is crucial for modulating related policies.

Objective: This study aimed to analyze stroke status in China from the 2023 Global Burden of Disease study.

Methods: The 2023 Global Burden of Disease study was the source of the estimation of stroke burden. The data of deaths, incidence, prevalence, disability-adjusted life years (DALYs), years lived with disability (YLDs), and years of life lost attributable to stroke and its subtypes from 1990 to 2023 were analyzed. The mortality-to-incidence ratio was calculated to quantify the proportion of incident stroke cases that result in deaths. DALY decomposition analysis was applied to quantify how stroke burden has shifted between mortality and long-term disability over time. Sex-specific characteristics were also compared.

Results: From 1990 to 2023, absolute stroke incidence (1.87-4.23 million) and prevalence (11.56-26.68 million) increased substantially, with ischemic stroke as the primary driver. Age-standardized incidence declined from 246.15 to 192.19 per 100,000 populations, while mortality significantly reduced from 302.72 to 92.83 per 100,000 populations. DALYs and years of life lost decreased 9.3% and 11%, respectively, while YLDs increased more than double. Mortality-to-incidence ratio fell for all subtypes except subarachnoid hemorrhage, indicating improved survival. Male individuals bore a heavier burden in total stroke and ischemic stroke, while subarachnoid hemorrhage showed female predominance in prevalence and YLDs. DALY decomposition revealed that YLDs contribution rose from 5.8% to 14.5%.

Conclusions: From 1990 to 2023, despite declining age-standardized rates, stroke burden in China grew due to ischemic stroke and population changes. The shifting burden toward disability and persistent sex disparities highlights the need for tailored prevention, rehabilitation, and gender-specific interventions.

(Online J Public Health Inform 2026;18:e86225) doi:10.2196/86225

KEYWORDS

stroke; ischemic stroke; intracerebral hemorrhage; subarachnoid hemorrhage; burden of disease

Introduction

Stroke, encompassing ischemic stroke (IS), intracerebral hemorrhage, and subarachnoid hemorrhage (SAH), constitutes one of the leading causes of mortality, disability, and socioeconomic burden globally [1]. Characterized by acute onset, high morbidity, and long-term sequelae, stroke imposes profound challenges to health care systems and public health strategies worldwide [2]. In China, demographic transitions such as population aging; epidemiological shifts such as rising prevalence of hypertension, diabetes, and obesity; and changes in lifestyle factors have collectively shaped the landscape of stroke burden over the past three decades [3-5]. Understanding the temporal trends, subtype-specific patterns, and sex or age disparities in stroke incidence, mortality, prevalence, and disability-adjusted life years (DALYs) is critical for optimizing resource allocation, guiding preventive interventions, and evaluating the effectiveness of health care policies.

Over the past decades, China has experienced remarkable advancements in stroke prevention, acute care, and rehabilitation [4]. However, the absolute burden of stroke remains substantial due to population growth and aging [6]. Previous studies based on earlier iterations of the Global Burden of Disease (GBD) study have reported declining age-standardized mortality rates (ASMRs) for stroke in China, but the trends in incidence, prevalence, and DALYs—particularly across different stroke subtypes—remain incompletely characterized [7]. Moreover, the relative contributions of mortality-related burden (years of life lost [YLLs]) and disability-related burden (years lived with disability [YLDs]) to total DALYs have not been fully elucidated, which is essential for prioritizing interventions targeting either acute survival or long-term functional recovery.

XSL•FO
RenderX

The GBD Study 2021 provides the latest and most comprehensive estimates of global disease burden, incorporating updated data sources, improved statistical modeling, and expanded geographic coverage [8]. Leveraging this dataset, this study aims to (1) analyze the temporal trends in incidence, mortality, prevalence, DALYs, YLLs, and YLDs of total stroke and its major subtypes in China from 1990 to 2023; (2) explore sex-specific disparities in stroke burden; (3) evaluate changes in mortality-to-incidence ratio (MIR) to assess survival improvements; and (4) decompose DALYs to quantify the shifting contributions of YLLs and YLDs over time.

Methods

Data Source

This study was based on the GBD Study 2023, the latest version of GBD report [9]. Definitions of the diseases included by the GBD study are based on the International Classification of Diseases [10]. The details of data source and methodology of the GBD study can be accessible at the official website [11]. Data were extracted from the Global Health Data Exchange database established by the Institute for Health Metrics and Evaluation of America. Options in the search panel on this website were set up as follows: (1) GBD estimate: causes of death or injury; (2) measures: deaths, incidence, prevalence, DALYs, YLDs, and YLLs; (3) metric: number; (4) cause: stroke (IS, intracerebral hemorrhage, and SAH); (5) location: China; (6) age: all ages; (7) sex: both, male and female; and (8) year: years from 1990 to 2023. In addition, the age-standardized rate for incidence and death in 1990, 2021, and 2023 were also extracted.

Data Processing

Data were presented as mean (95% uncertainty interval). The uncertainty levels were determined by the GBD modeling process as previously reported [12]. Values of lower and upper uncertainty can be extracted from the data source.

Burden of stroke, IS, intracerebral hemorrhage, and SAH were first analyzed. Sex-specific trends of the disease burden were consequently displayed respectively. The MIR was further calculated to quantify the proportion of incident stroke cases that result in death, avoiding complex population denominator dependencies. Lower MIR represents better survival (reflecting improved acute care, rehabilitation, or treatment efficacy).

DALY decomposition analysis was applied to decompose DALYs into its 2 core components—YLLs (mortality-related burden) and YLDs (disability-related burden)—to quantify how stroke burden has shifted between mortality and long-term disability over time.

Ethical Considerations

This study did not include human or animal participants; therefore, ethics approval was not required for this study. For this type of study, formal consent was also not required.

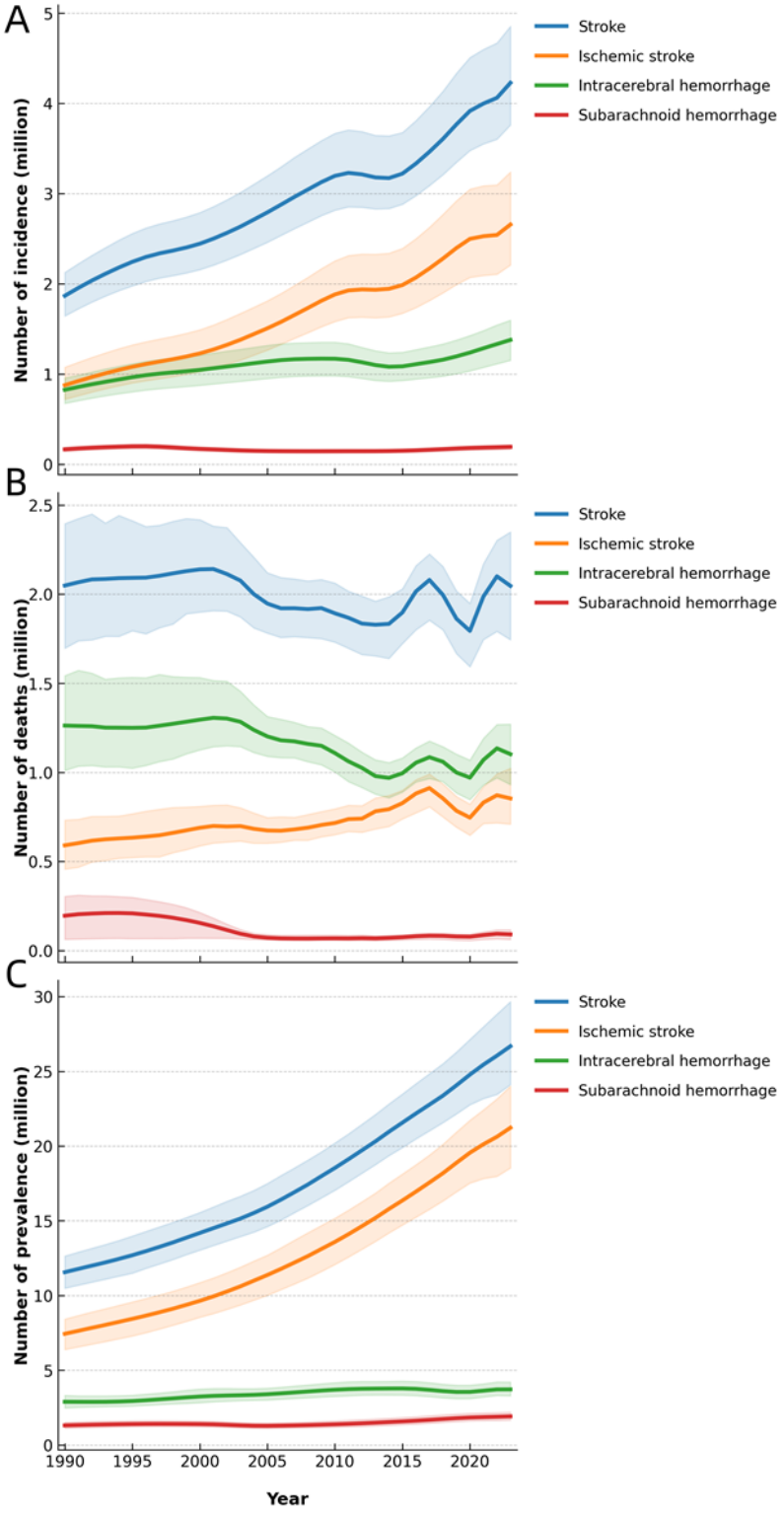
Results

Disease Burden of Stroke in China From 1990 to 2023

The absolute number of new stroke cases increased from 1.87 million in 1990 to 4.23 million in 2023 (+126%; Figure 1A). The age-standardized incidence rate (ASIR) declined from 246.15 per 100,000 population in 1990 to 192.19 per 100,000 in 2023 (–22%; Table 1). Total stroke deaths remained little changed from 1990 to in 2023 (Figure 1B), but the ASMR declined from 302.72 per 100,000 in 1990 to 92.83 per 100,000 in 2023 (–69%; Table 1). The total number of stroke patients dramatically increased from 11.56 million in 1990 to 26.68 million in 2023 (+131%; Figure 1C). From 1990 to 2023, the DALYs and YLLs decreased, but YLDs largely increased from 22.14 million to 51.33 million (Figure 2A-2C).

XSL•FO
RenderX

Figure 1. The trends of incidence (A), deaths (B), and prevalence (C) for stroke, ischemic stroke, intracerebral hemorrhage, and subarachnoid hemorrhage from 1990 to 2023.



XSL•FO
RenderX

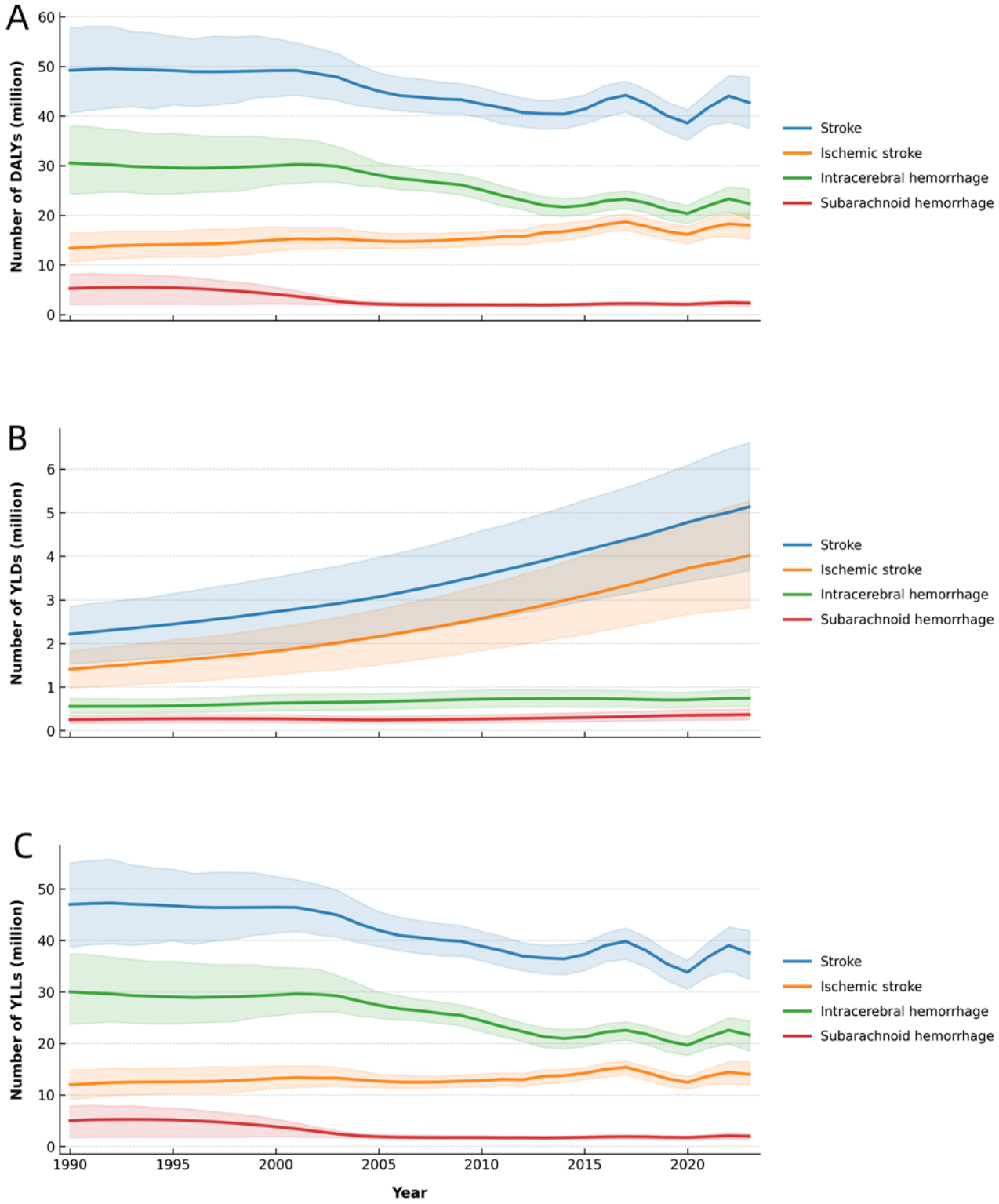
Table . The age-standardized rate of incidence and death for stroke, ischemic stroke, intracerebral hemorrhage, and subarachnoid in 1990, 2021, and 2023.

	Age-standardized incidence rate per 100,000 (95% UI)[a]			Age-standardized death rate per 100,000 (95% UI)		
	1990	2021	2023	1990	2021	2023
Stroke	246.15 (216.42-280.69)	193.54 (172.89-221.59)	192.19 (172.08-219.81)	302.72 (252.74-357.05)	97.48 (85.00-106.85)	92.83 (79.01-106.61)
Ischemic stroke	112.33 (93.74-135.06)	119.89 (100.86-143.77)	118.40 (100.04-141.31)	94.12 (74.19-116.98)	41.33 (35.21-45.99)	39.02 (32.35-46.75)
Intracerebral hemorrhage	114.44 (94.13-132.56)	64.01 (53.96-73.80)	64.16 (54.47-74.08)	181.68 (148.08-225.91)	51.96 (44.58-58.00)	49.72 (42.03-57.36)
Subarachnoid hemorrhage	19.39 (16.44-22.37)	9.63 (8.38-10.93)	9.63 (8.38-10.91)	5.69 (4.52-6.64)	4.19 (2.99-5.13)	4.10 (2.81-5.27)

[a]UI: uncertainty interval.

XSL•FO
RenderX

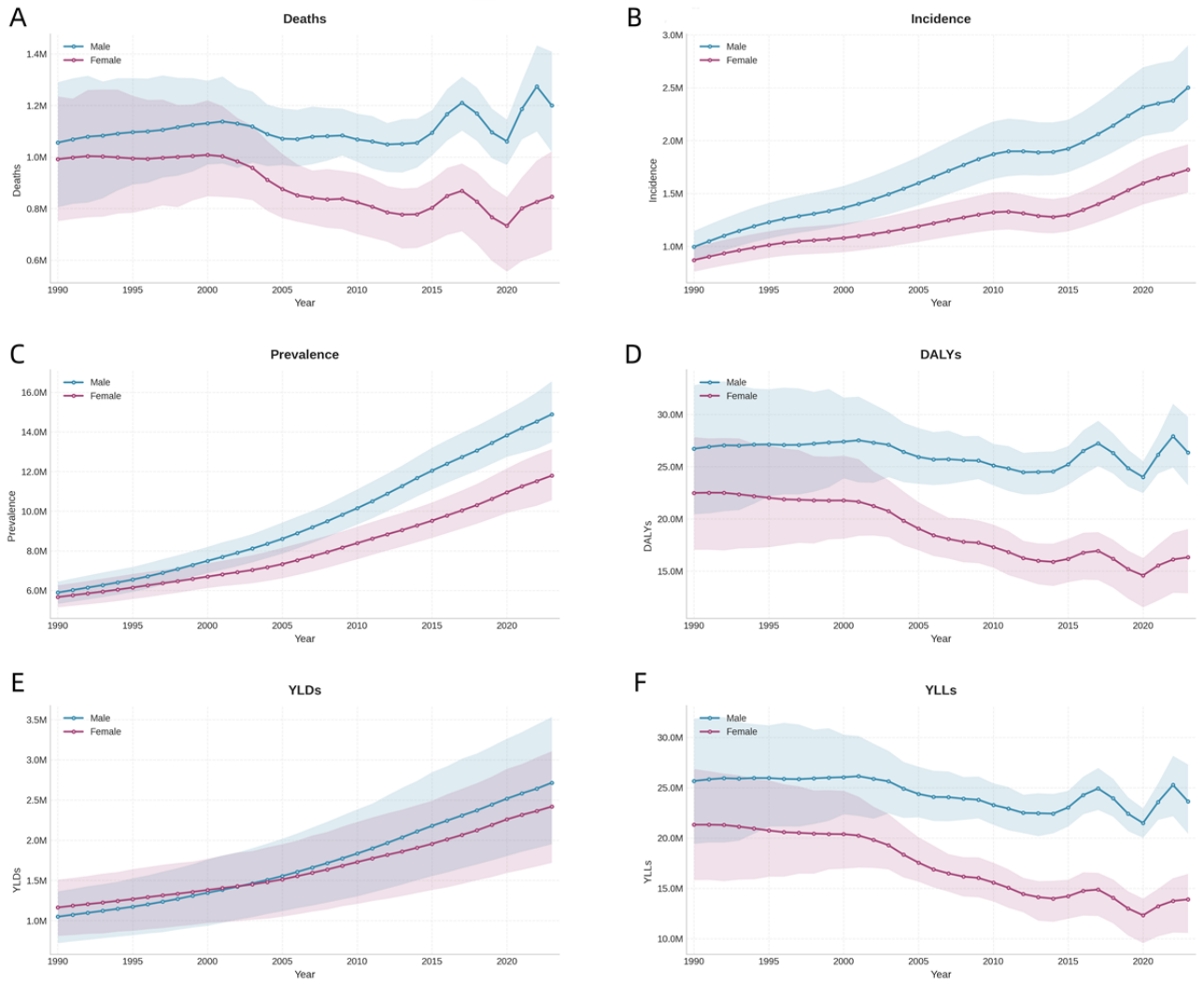
Figure 2. The trends of disability-adjusted life years (DALYs; **A**), years lived with disability (YLDs; **B**), and years of life lost (YLLs; **C**) for stroke, ischemic stroke, intracerebral hemorrhage, and subarachnoid hemorrhage from 1990 to 2023.



A consistent male predominance in stroke burden was observed across all metrics throughout the study period, though the magnitude and trajectory of the gender gap varied by outcome (Figure 3A-3F). From 1990 to 2023, male stroke deaths slightly increased, while female deaths reduced (Figure 3A). Although the incidence, prevalence, and YLDs of both male and female increased, male individuals showed a larger increase in amplitude (Figure 3B, 3C, and 3E). DALYs and YLLs of male individuals slightly decreased, while female individuals had a significant reduction (Figure 3D and 3F).

XSL•FO
RenderX

Figure 3. Gender-specific trends of stroke burden in China. DALY: disability-adjusted life year; YLD: years lived with disability; YLL: years of life lost.



Disease Burden of IS in China From 1990 to 2023

The absolute number of new IS cases rose substantially from 0.88 million in 1990 to 2.66 million in 2023 (+202%; Figure 1A). Align with the growth in absolute incidence, the ASIR of IS also increased from 112.33 per 100,000 population in 1990 to 118.40 per 100,000 in 2023 (Table 1). Total IS deaths exhibited a rise over the study period (Figure 1B), but a significant ASMR plummeted from 94.12 per 100,000 in 1990 to 39.02 per 100,000 in 2023 (Table 1). The total number of people living with IS surged from 7.44 million in 1990 to 21.23 million in 2023 (Figure 1C). DALY-related metrics of IS all increased from 1990 to 2023, with YLDs showing a marked increase (Figure 2A-2C).

Male predominance in IS burden was consistently increased across all metrics, though the size and trajectory of the gender gap differed by outcome (Figure 4A-4F).

XSL•FO
RenderX

Figure 4. Gender-specific trends of ischemic stroke burden in China. DALY: disability-adjusted life year; YLD: years lived with disability; YLL: years of life lost.

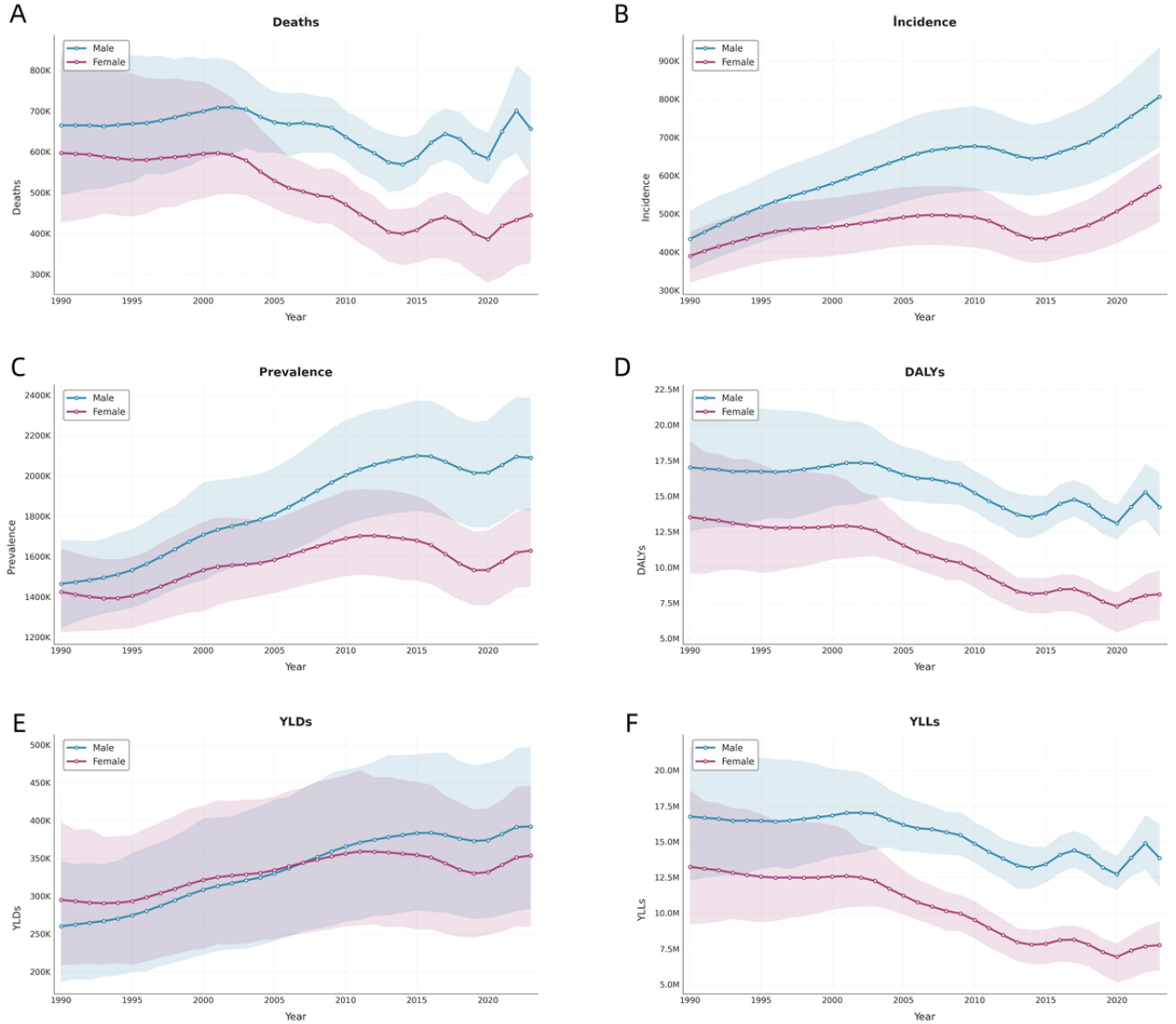
Disease Burden of Intracerebral Hemorrhage in China From 1990 to 2023

The incidence of intracerebral hemorrhage increased moderately from 0.83 million in 1990 to 1.38 million in 2023 (+66%; Figure 1A). In contrast to the rise in absolute incidence, the ASIR of intracerebral hemorrhage declined notably, falling from 114.44 per 100,000 population in 1990 to 64.16 per 100,000 in 2023 (−44%; Table 1). Total intracerebral hemorrhage deaths slightly decreased from 1990 to 2023 (Figure 1B), but ASMR showed a dramatic downward trend—dropping from 181.68 per 100,000 in 1990 to 49.72 per 100,000 in 2023 (–73%; Table 1). Benefiting from improved survival alongside mild incidence growth, the prevalence number rose from 2.89 million in 1990 to 3.72 million in 2023 (+29%; Figure 1C). Among DALY-related metrics, intracerebral hemorrhage DALYs and YLLs (mortality burden) both decreased, while YLDs (disability burden) increased moderately (Figure 2A-2C).

Male predominance in intracerebral hemorrhage burden was consistent across all core metrics except YLDs (Figure 5A-5F). The YLDs of intracerebral hemorrhage in female individuals were higher than that in male individuals before 2007, while YLDs in male individuals have surpassed since 2008 (Figure 5E).

XSL•FO
RenderX

Figure 5. Gender-specific trends of intracerebral hemorrhage burden in China. DALY: disability-adjusted life year; YLD: years lived with disability; YLL: years of life lost.



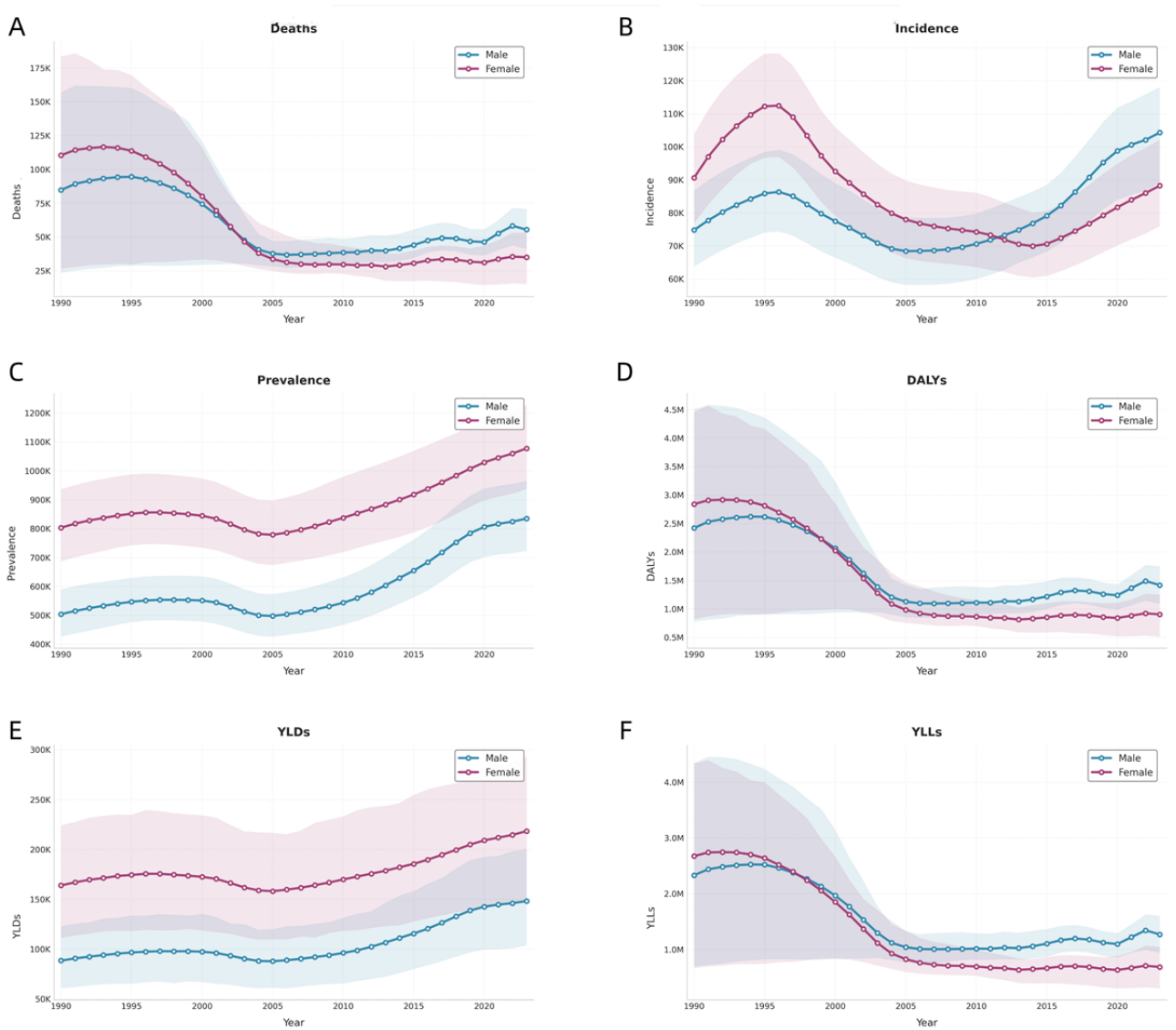
Disease Burden of SAH in China From 1990 to 2023

The absolute number of new SAH cases increased slightly from 0.17 million in 1990 to 0.19 million in 2023 (Figure 1A). In contrast to the rise in absolute incidence, the ASIR of SAH showed a downward trend, falling from 19.39 per 100,000 population in 1990 to 9.63 per 100,000 in 2023 (−50%; Table 1). Total SAH deaths increased from 0.20 million in 1990 to 0.91 million in 2023 (+355%; Figure 1B), while the ASMR dropped slightly from 5.69 per 100,000 in 1990 to 4.10 per 100,000 in 2023 (−28%; Table 1). The total number of people living with SAH rose from 1.31 million in 1990 to 1.91 million in 2023 (Figure 1C). Among DALY-related metrics, SAH DALYs and YLLs both decreased notably, while YLDs increased moderately, expanding from 0.25 million in 1990 to 0.37 million in 2023 (Figure 2A-2C).

Unlike other subtypes of stroke, the prevalence and YLDs of female were larger than that of male individuals during the study period (Figure 6C and 6E). The incidence, deaths, DALYs, and YLLs of female individuals were once higher than that of male individuals but have been surpassed (Figure 6A, 6B, 6D, and 6F).

XSL•FO
RenderX

Figure 6. Gender-specific trends of subarachnoid hemorrhage burden in China. DALY: disability-adjusted life year; YLD: years lived with disability; YLL: years of life lost.



Analysis of MIR

All stroke types except SAH showed reduced MIR from 1990 to 2023 (Table 2). Intracerebral hemorrhage had the highest MIR in both years, while IS showed a largest reduction rate (Table 2). In 1990, male individuals had higher MIR than female individuals for all subtypes, while the MIR in 2023 was opposite except SAH (Table 2).

XSL•FO
RenderX

Table . Mortality-to-incidence ratio of stroke and its subtypes in 1990 and 2023.

	Total (95% UI)[a]	Male (95% UI)	Female (95% UI)
Stroke			
1990	1.23 (1.07 - 1.42)	1.27 (1.09 - 1.48)	1.19 (1.02 - 1.39)
2023	0.48 (0.42 - 0.56)	0.46 (0.40 - 0.53)	0.51 (0.44 - 0.59)
Ischemic stroke			
1990	0.84 (0.69 - 1.03)	0.87 (0.71 - 1.07)	0.81 (0.65 - 1.01)
2023	0.33 (0.28 - 0.40)	0.31 (0.26 - 0.38)	0.35 (0.29 - 0.43)
Intracerebral hemorrhage			
1990	1.59 (1.34 - 1.89)	1.64 (1.37 - 1.96)	1.54 (1.28 - 1.85)
2023	0.77 (0.66 - 0.91)	0.74 (0.63 - 0.88)	0.80 (0.68 - 0.95)
Subarachnoid hemorrhage			
1990	0.29 (0.23 - 0.37)	0.31 (0.24 - 0.40)	0.28 (0.21 - 0.36)
2023	0.43 (0.33 - 0.57)	0.45 (0.34 - 0.60)	0.41 (0.31 - 0.55)

[a]UI: uncertainty interval.

Decomposition Analysis of Stroke DALYs

Decomposition of stroke DALYs is shown in Table 3. For the total population, DALY composition shifted uniformly from mortality to disability between 1990 and 2023: YLLs contribution declined across all subtypes while YLDs contribution nearly tripled for total stroke (5.8%-14.5%) and doubled for SAH (8.3%-17.1%), with IS becoming the most disability-heavy subtype. Male individuals consistently had higher YLLs contribution and lower YLDs contribution than female individuals. Female individuals, by contrast, had the steepest YLLs decline and fastest YLDs growth, driving the total population's disability shift and widening the sex gap in YLDs contribution from 2% to 6.7% for total stroke. Intracerebral hemorrhage remained mortality-dominant (YLLs ≥95.5% in 2023) across all groups, though female individuals still had slightly higher YLDs contribution than male individuals for this subtype.

Table . Decomposition of stroke disability-adjusted life years into years of life lost (YLLs) and years lived with disability (YLDs) by sex in 1990 and 2023.

	Total population		Male		Female	
	YLLs contribution, % (95% UI)[a]	YLDs contribution, % (95% UI)	YLLs contribution, % (95% UI)	YLDs contribution, % (95% UI)	YLLs contribution, % (95% UI)	YLDs contribution, % (95% UI)
Stroke						
1990	94.2 (93.5 - 95.0)	5.8 (5.0 - 6.5)	95.1 (94.3 - 95.8)	4.9 (4.2 - 5.7)	93.1 (92.2 - 94.0)	6.9 (6.0 - 7.8)
2023	85.5 (84.3 - 86.7)	14.5 (13.3 - 15.7)	88.8 (87.5 - 90.0)	11.2 (10.0 - 12.5)	82.1 (80.6 - 83.6)	17.9 (16.4 - 19.4)
Ischemic stroke						
1990	92.6 (91.7 - 93.5)	7.4 (6.5 - 8.3)	93.5 (92.5 - 94.4)	6.5 (5.6 - 7.5)	91.5 (90.3 - 92.7)	8.5 (7.3 - 9.7)
2023	77 (75.6 - 78.4)	23(21.6 - 24.4)	79.7 (78.0 - 81.4)	20.3 (18.6 - 22.0)	74.1 (72.2 - 76.0)	25.9 (24.0 - 27.8)
Intracerebral hemorrhage						
1990	97.1 (96.4 - 97.8)	2.9 (2.2 - 3.6)	97.5 (96.8 - 98.1)	2.5 (1.9 - 3.2)	96.7 (95.9 - 97.5)	3.3 (2.5 - 4.1)
2023	96.2 (95.4 - 97.0)	3.8 (3.0 - 4.6)	96.8 (96.0 - 97.5)	3.2 (2.5 - 4.0)	95.5 (94.5 - 96.5)	4.5 (3.5 - 5.5)
Subarachnoid hemorrhage						
1990	91.7 (90.2 - 93.1)	8.3 (6.9 - 9.8)	92.4 (90.7 - 94.0)	7.6 (6.0 - 9.3)	90.9 (89.0 - 92.7)	9.1 (7.3 - 11.0)
2023	82.9 (80.8 - 85.0)	17.1 (15.0 - 19.2)	`85.3 (82.8 - 87.8)	14.7 (12.2 - 17.2)	80.4 (77.8 - 83.0)	19.6 (17.0 - 22.2)

[a]UI: uncertainty interval.

Discussion

In general, the burden of stroke increased from 1990 to 2023 in China, despite the decline of ASIR and mortality rate. Our study showed that the overall deaths, incidence, and prevalence increased by 11.8%, 38.2%, and 130.8%, respectively, while DALYs and YLLs decreased by 9.3% and 11%, respectively. Our results were based on the estimated number, which is

XSL•FO
RenderX

different from the previous study [13]. The reason why we chose the estimated number as the metric is that the increasing number of stroke patients is generally associated with increasing financial burden and consumption of medical resources. Thus, the increase of stroke-related burden should be attached much more attention.

Specifically, the increase mainly came from IS, though the burden of intracerebral hemorrhage and subarachnoid also slightly increased. IS, primarily caused by embolism and arterial atherosclerotic disease, often presents as sudden-onset unilateral limb paralysis, facial asymmetry, abnormal speaking, confusion, changes of consciousness, unbalance or incoordination, and visual impairment [14,15]. As well recognized, hypertension, smoking, hyperlipemia, diabetes, transient ischemic attacks, and atrial fibrillation are the common risk factors for IS. A previous study based on 15 nationwide surveys pointed out that the prevalence of hypertension in China was 18% to 44.7% [16]. A secondary study from a national cross-sectional study showed that 2.31% of the enrolled 0.7 million participants had atrial fibrillation [17]. Moreover, metabolic factors were more severe, with an estimated morbidity of 31% and 38% for hyperlipemia and diabetes, respectively [18,19]. In addition, as the largest producer and consumer of cigarettes in the world, the smoking rate was estimated to be over 20% in China [20]. All these risks contributed to the increased incidence of IS. However, the mortality of IS has declined significantly over the past 30 years.

Progresses have been made to improve the survival of IS, including the treatment and prevention strategies. Of most importance, initiation of thrombolysis within 4.5 hours and endovascular thrombectomy within 24 hours after symptom onset are considered to improve neural outcomes and are recommended by the guidelines [21-23]. The composition of thrombolytic drugs (Tenecteplase, etc) and anesthesia methods (procedural sedation and general anesthesia) have also been investigated and applied to acute IS [24,25]. The whole cycle management of IS advocates early detection and reperfusion, and a multidisciplinary care in hospital and out of hospital. In this study, the MIR significantly decreased from 1990 to 2023, indicating great improvement in stroke care. All these advances, consequently, augment the burden of DALYs and YLDs. Our results of DALYs and YLDs follow this rule. Another vital policy of China is to establish stroke centers, where the stroke patients can be treated as soon as possible in a standardized procedure [26]. Stroke centers are correlated to better outcomes of patients with acute IS [27,28].

These 34-year data on stroke burden in China reveal notable sex-specific patterns across total stroke and its subtypes, driven by biological differences, gendered risk factor exposure, and health care access, which warrant targeted public health attention. A notable point is the widening gaps between male and female individuals in the burden of stroke—a disease highly related to lifestyle [29,30]. In China, male individuals had higher rates of smoking and heavy alcohol use [31,32].

When interpreting the results, we should first understand the intrinsic limitations of the GBD study. As described, the estimated disease burden is based on the availability of original data and calculation models, which makes it difficult to fully represent the actual situation [12]. The inherent uncertainties in GBD estimations and potential reliance on limited or low-quality data for certain inputs should also be noticed. Nevertheless, the GBD study is the most reliable tool to evaluate global health issues. Secondly, the GBD study does not provide province-level data, which is important to understand the regional characteristics. Finally, the study does not account for the impact of potential confounding factors on stroke burden trends, such as the contributions of specific modifiable risk factors (eg, regional variations in hypertension control rates, smoking prevalence, or dietary patterns) or nonmodifiable factors (eg, genetic susceptibility) to the observed trends.

In conclusion, from 1990 to 2023, stroke burden in China saw growing absolute numbers despite declining ASIRs and mortality rates, driven primarily by IS. The burden shifted toward disability as evidenced by the decreased MIR and increased YLDs' contribution to DALYs. Male individuals faced heavier stroke burden. These trends call for tailored strategies: IS prevention, enhanced rehabilitation for disability, and gender-specific interventions alongside sustained investment in stroke care systems.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

References

1. Feigin VL, Brainin M, Norrving B, et al. World Stroke Organization: Global Stroke Fact Sheet 2025. Int J Stroke 2025 Feb;20(2):132-144. [doi: 10.1177/17474930241308142] [Medline: 39635884]
2. Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. Circulation 2019 Mar 5;139(10):e56-e528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
3. Wang W, Jiang B, Sun H, et al. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults. Circulation 2017 Feb 21;135(8):759-771. [doi: 10.1161/CIRCULATIONAHA.116.025250] [Medline: 28052979]

XSL•FO
RenderX

4. Tu WJ, Wang LD, Special Writing Group of China Stroke Surveillance Report. China stroke surveillance report 2021. Mil Med Res 2023 Jul 19;10(1):33. [doi: 10.1186/s40779-023-00463-x] [Medline: 37468952]
5. Guan T, Ma J, Li M, et al. Rapid transitions in the epidemiology of stroke and its risk factors in China from 2002 to 2013. Neurology 2017 Jul 4;89(1):53-61. [doi: 10.1212/WNL.0000000000004056] [Medline: 28566547]
6. Zhou M, Wang H, Zhu J, et al. Cause-specific mortality for 240 causes in China during 1990-2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. Lancet 2016 Jan 16;387(10015):251-272. [doi: 10.1016/S0140-6736(15)00551-6] [Medline: 26510778]
7. Yang G, Wang Y, Zeng Y, et al. Rapid health transition in China, 1990-2010: findings from the Global Burden of Disease Study 2010. Lancet 2013 Jun 8;381(9882):1987-2015. [doi: 10.1016/S0140-6736(13)61097-1] [Medline: 23746901]
8. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2100-2132. [doi: 10.1016/S0140-6736(24)00367-2] [Medline: 38582094]
9. GBD results. Institute for Health Metrics and Evaluation. URL: https://vizhub.healthdata.org/gbd-results/ [accessed 2026-05-26]
10. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2133-2161. [doi: 10.1016/S0140-6736(24)00757-8] [Medline: 38642570]
11. Global Burden of Disease (GBD). Institute for Health Metrics and Evaluation. URL: https://www.healthdata.org/research-analysis/about-gbd [accessed 2026-01-20]
12. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
13. Wang YJ, Li ZX, Gu HQ, et al. China Stroke Statistics: an update on the 2019 report from the National Center for Healthcare Quality Management in Neurological Diseases, China National Clinical Research Center for Neurological Diseases, the Chinese Stroke Association, National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention and Institute for Global Neuroscience and Stroke Collaborations. Stroke Vasc Neurol 2022 Oct;7(5):415-450. [doi: 10.1136/svn-2021-001374] [Medline: 35443985]
14. Feske SK. Ischemic stroke. Am J Med 2021 Dec;134(12):1457-1464. [doi: 10.1016/j.amjmed.2021.07.027] [Medline: 34454905]
15. Walter K. What is acute ischemic stroke? JAMA 2022 Mar 1;327(9):885. [doi: 10.1001/jama.2022.1420] [Medline: 35230392]
16. Yin R, Yin L, Li L, et al. Hypertension in China: burdens, guidelines and policy responses: a state-of-the-art review. J Hum Hypertens 2022 Feb;36(2):126-134. [doi: 10.1038/s41371-021-00570-z] [Medline: 34215840]
17. Wang X, Fu Q, Song F, et al. Prevalence of atrial fibrillation in different socioeconomic regions of China and its association with stroke: results from a national stroke screening survey. Int J Cardiol 2018 Nov 15;271:92-97. [doi: 10.1016/j.ijcard.2018.05.131] [Medline: 29885822]
18. Wang L, Peng W, Zhao Z, et al. Prevalence and treatment of diabetes in China, 2013-2018. JAMA 2021 Dec 28;326(24):2498-2506. [doi: 10.1001/jama.2021.22208] [Medline: 34962526]
19. Peng K, Cai W, Liu X, et al. Trends of hypercholesterolemia change in Shenzhen, China during 1997-2018. Front Public Health 2022;10:887065. [doi: 10.3389/fpubh.2022.887065] [Medline: 35586010]
20. Yuan L, Liu P, Zhao Z, Wei Z, Liu L, Sun J. Cross-sectional survey on cigarette smoking in Chinese high-income areas. BMJ Open 2022 Apr 29;12(4):e056209. [doi: 10.1136/bmjopen-2021-056209] [Medline: 35487748]
21. Mendelson SJ, Prabhakaran S. Diagnosis and management of transient ischemic attack and acute ischemic stroke: a review. JAMA 2021 Mar 16;325(11):1088-1098. [doi: 10.1001/jama.2020.26867] [Medline: 33724327]
22. Chen HS, Cui Y, Zhou ZH, et al. Dual antiplatelet therapy vs alteplase for patients with minor nondisabling acute ischemic stroke: the ARAMIS randomized clinical trial. JAMA 2023 Jun 27;329(24):2135-2144. [doi: 10.1001/jama.2023.7827] [Medline: 37367978]
23. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019 Dec;50(12):e344-e418. [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]
24. Tsivgoulis G, Katsanos AH, Sandset EC, et al. Thrombolysis for acute ischaemic stroke: current status and future perspectives. Lancet Neurol 2023 May;22(5):418-429. [doi: 10.1016/S1474-4422(22)00519-1] [Medline: 36907201]
25. Campbell D, Butler E, Campbell RB, Ho J, Barber PA. General anesthesia compared with non-GA in endovascular thrombectomy for ischemic stroke: a systematic review and meta-analysis of randomized controlled trials. Neurology 2023 Apr 18;100(16):e1655-e1663. [doi: 10.1212/WNL.0000000000207066] [Medline: 36797071]
26. Chao BH, Yan F, Hua Y, et al. Stroke prevention and control system in China: CSPPC-Stroke program. Int J Stroke 2021 Apr;16(3):265-272. [doi: 10.1177/1747493020913557] [Medline: 32223541]

XSL•FO
RenderX

27. Shen Y, Chao BH, Cao L, Tu WJ, Wang LD. Stroke center care and outcome: results from the CSPPC Stroke program. Transl Stroke Res 2020 Jun;11(3):377-386. [doi: 10.1007/s12975-019-00727-6] [Medline: 31494833]
28. Man S, Xian Y, Holmes DN, et al. Association between thrombolytic door-to-needle time and 1-year mortality and readmission in patients with acute ischemic stroke. JAMA 2020 Jun 2;323(21):2170-2184. [doi: 10.1001/jama.2020.5697] [Medline: 32484532]
29. Rutten-Jacobs LC, Larsson SC, Malik R, et al. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. BMJ 2018 Oct 24;363:k4168. [doi: 10.1136/bmj.k4168] [Medline: 30355576]
30. Ma Q, Li R, Wang L, et al. Temporal trend and attributable risk factors of stroke burden in China, 1990-2019: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2021 Dec;6(12):e897-e906. [doi: 10.1016/S2468-2667(21)00228-0] [Medline: 34838196]
31. Ding L, Song B, Wu C, et al. Alcohol use in China: unrecorded and recorded bai jiu in three rural regions. Int J Environ Res Public Health 2021 Dec 30;19(1):405. [doi: 10.3390/ijerph19010405] [Medline: 35010665]
32. Chen W, Xia C, Zheng R, et al. Disparities by province, age, and sex in site-specific cancer burden attributable to 23 potentially modifiable risk factors in China: a comparative risk assessment. Lancet Glob Health 2019 Feb;7(2):e257-e269. [doi: 10.1016/S2214-109X(18)30488-1] [Medline: 30683243]

Abbreviations

ASIR: age-standardized incidence rate
ASMR: age-standardized mortality rate
DALY: disability-adjusted life year
GBD: Global Burden of Disease
IS: ischemic stroke
MIR: mortality-to-incidence ratio
SAH: subarachnoid hemorrhage
YLDs: years lived with disability
YLLs: years of life lost

Edited by E Mensah; submitted 21.Oct.2025; peer-reviewed by SY Guan, Y Li; revised version received 04.Mar.2026; accepted 16.May.2026; published 22.Jun.2026.

Please cite as:
Fu L, Jiang W, Peng Y, Zhang X, Zhou R
Burden of Stroke in China From 1990 to 2023: Analysis From the Global Burden of Disease Study 2023
Online J Public Health Inform 2026;18:e86225
URL: https://ojphi.jmir.org/2026/1/e86225
doi:10.2196/86225

© Lijuan Fu, Wencai Jiang, Yanhua Peng, Xianjie Zhang, Rui Zhou. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 22.Jun.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Building Enhanced Public Health Data Systems With a Situational Awareness and Learning Tool: Focus Group Study

Cole Brokamp[1,2], PhD; Carson S Hartlage[1], BS; Tiffany Mattingly[3], MS; Pierce Kuhnell[2], MS; Andrew Vancil[2], MS; Andrew F Beck[1,4,5,6,7,8], MD, MPH; David Hartley[1,5], PhD

[1]Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, OH, United States

[2]Division of Biostatistics and Epidemiology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH, United States

[3]The Health Collaborative, Cincinnati, OH, United States

[4]Division of General and Community Pediatrics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[5]James M Anderson Center for Health Systems Excellence, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[6]Division of Hospital Medicine, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[7]Michael A. Fisher Child Health Equity Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[8]Office of Population Health, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

Corresponding Author:
Cole Brokamp, PhD
Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, OH, United States

Abstract

Background: Situational awareness is the accurate and timely perception of factors in the environment, comprehension of their meanings, and projection of their future states.

Objective: We aimed to develop a cloud-based Situational Awareness and Learning Tool (SALT) that generates near-real-time analytic content from multimodal health care, government, community, and environmental data, enabling public health and hospital professionals to make informed decisions during complex population health challenges.

Methods: Several focus groups were conducted with representatives from local health departments, hospitals, and emergency agencies. The first round identified data needs and requirements to inform SALT's design. SALT was developed as a secure, cloud-based platform featuring automated deployment, role-based access, and version-controlled content publishing. The second round of focus groups evaluated the SALT prototype's utility and gathered feedback for improvements.

Results: Participants highlighted the need for integrated data from multiple sources, tailored dashboards for specific audiences, and legal frameworks to guide timely data sharing. SALT met these requirements by providing interactive visuals, secure access levels, and a collaborative content management system. The second focus groups affirmed SALT's effectiveness in enhancing decision-making and strategic planning, suggesting enhancements such as clearer data labeling, expanded data coverage, and forecasting capabilities.

Conclusions: SALT addresses limitations exposed by the COVID-19 pandemic in public health data systems by offering a scalable platform for data sharing, rapid analysis, and situational awareness. It fulfills user needs for integrated, timely data, and customized analytic products. SALT represents a viable solution for enhancing public health data systems in preparation for future pandemics and other complex, multisector population health challenges.

(Online J Public Health Inform 2026;18:e77379) doi:10.2196/77379

KEYWORDS

data systems; public health informatics; evaluation study; public health; infectious disease

Introduction

Public health data systems are essential for monitoring, preventing, and responding to infectious diseases, especially those with pandemic potential. To be effective, systems should support comprehensive situational awareness: accurate and timely perception of factors affecting public health and well-being, comprehension of their meanings, and projection of key factors into the future [1]. The COVID-19 pandemic exposed the limitations and gaps in existing data systems at local, state, national, and global levels for providing adequate situational awareness. For example, in the early waves of

XSL•FO
RenderX

COVID-19, many hospitals and public health agencies faced challenges in collecting, integrating, analyzing, and sharing timely and accurate data on testing, cases, hospitalizations, deaths, vaccinations, and other indicators of COVID-19 transmission and impact [2]. These challenges are broad and can be related to (1) technology and interoperability of systems, (2) access and privacy concerns, and (3) a lack of reproducible and scientific data reporting pipelines [3-5]. Such difficulties persist, placing our public health and health care systems at a significant disadvantage [6]. Moreover, many data systems lack the capacity to incorporate information on the social and environmental determinants of health. Indeed, factors such as poverty, housing, air quality, and mobility affect the vulnerability and resilience of populations to COVID-19 and other health threats [7]. As a result, public health decision-makers and partners often lack the situational awareness and learning tools needed to effectively plan, implement, and evaluate evidence-based and equity-oriented interventions to mitigate and prevent COVID-19 and future pandemics [8].

Recent work further emphasizes that these limitations extend beyond outbreak surveillance alone. Inadequate integration of social, demographic, and environmental determinants of health, including housing conditions, poverty, air quality, and mobility patterns, has constrained the ability of public health agencies to identify vulnerable populations and deploy equity-oriented interventions in real time [5,7]. National reviews of pandemic data infrastructures highlight that delays and inconsistencies in reporting materially affected operational decision-making and resource allocation, reinforcing calls for modernized, interoperable, and policy-aligned public health data ecosystems [4-6].

In response to these calls, federal agencies and public health leaders have increasingly emphasized investments in data modernization, cloud-based infrastructure, and interoperable surveillance architectures capable of supporting near-real-time analytics across clinical, community, and environmental domains [6]. Automated, reproducible platforms that harmonize multimodal data sources and disseminate tailored analytic products through role-based permissions have therefore emerged as a central component of pandemic preparedness and disaster response strategies [6,8,9].

With this backdrop, our objective was to create a situational awareness and learning tool (SALT) capable of generating near-real-time analytic content using multimodal data. We also sought to make SALT as useful as possible for frontline public health and hospital professionals, optimized in ways that allowed them to make informed decisions tuned to on-the-ground conditions.

Methods

Overview

SALT development drew upon our experience in the southwest Ohio COVID-19 response beginning in February 2020. During that time, we worked collaboratively with major interest holders, including hospital systems, public health, and the congregate care community, to identify data needs and then rapidly build an ad hoc situational awareness system [6-8]. When the public health emergency ended in May 2023, local interest holders expressed the need to perpetuate and expand the functionality of the legacy system beyond COVID-19.

Design Focus Groups

A qualitative methodology [10,11] using focus groups was selected to capture nuanced, context-specific insights from stakeholders involved in the COVID-19 response. This approach facilitated interactive discussion, allowing participants to collectively reflect on data challenges and inform the user-centered design and refinement of SALT.

We conducted 2 iterative rounds of focus groups, one centered on design and the second on the evaluation of SALT, in collaboration with a nonprofit partner, The Health Collaborative. For the first round, we purposively [12] sampled leaders from various agencies and organizations involved in the pandemic response in counties within the Ohio Hospital Association's COVID-19 Zone 6, including city and county public health departments, hospitals, and emergency management agencies (EMAs), using targeted email outreach.

Focus groups were held virtually on a video conferencing platform (Microsoft Teams) and used a semistructured interview format, with open-ended questions and prompts to guide discussion. Sessions lasted approximately 60 minutes. Each focus group was attended by at least three members of The Health Collaborative, who moderated and took field notes throughout. Sessions were audio recorded and transcribed. Following each session, the research team collaboratively reviewed notes and transcripts to discuss emergent ideas and prioritize themes. Multiple authors reviewed the outputs from the focus groups to summarize responses to specific questions and interpret similarities across different group discussions. Data were collected until no new ideas were mentioned by participants.

The first round of focus groups occurred in November 2023 and consisted of 3 sessions. The goal of these sessions was to identify data elements, sources, and information of potential benefit to ongoing COVID-19 phases and future pandemic preparedness. A standard agenda and interview guide were used for all sessions (Multimedia Appendix 1). The questions focused on organizational data access, data-informed decisions, frustrations with available data, and visions for an ideal data tool during future emergencies.

Themes from these focus groups were used to inform the initial design and functionality of SALT, with particular attention to desired data types, visualization needs, and system usability. In parallel, we compiled a list of data sources mentioned by participants and assessed each for availability, accessibility, and quality as potential inputs for the tool.

Initial SALT Build

The first set of focus groups, and our experience during various phases of the COVID-19 pandemic [2,8,9,13-16], informed the initial SALT build. We designed the SALT architecture to automatically deploy onto a cloud infrastructure provider as a

XSL•FO
RenderX

virtual private cloud (VPC). VPC offers advantages over on-premises hosting, including security, backup and failover, and extensible compute capability. To ensure high availability, we deployed SALT in 2 geographical data center regions and with fully automatic database failover in the event of a vendor hardware failure. Cloud-based computing resources are abundantly available and are dynamically allocated to the rendering or serving of content based on real-time demand. Heavily accessed items automatically receive additional resources to maintain functionality.

Access levels for SALT products include “public” (no account required), “authenticated” (account required), and “authorized” (account with product-specific permissions required). Authentication is achieved through federation with InCommon, and accounts are created by a SALT administrator after agreement to SALT policies. Content is created by SALT publishers, who create the code-based content for SALT users to consume. Authorization for specific content is managed by the content publisher and audited by SALT administrators.

To enhance the efficiency, security, and integrity of content publishing, a detailed procedure has been established for publishers. Initially, potential publishers engage with SALT administrators to evaluate the compatibility of their content for inclusion. Upon approval, publishers are required to complete a brief survey and agree to an acceptable use policy available on the SALT landing page. Following this, SALT administrators configure a user account, marking the entry into the publishing workflow.

For integration into SALT, all code used to generate content is housed within a version-controlled Git repository. Contributors without existing version control are supported in setting up a new repository within the SALT organization on GitHub. For contributors with a preexisting repository, that repository is forked into the SALT organization to ensure that any updates follow the established review and integration protocols.

Publishers initiate content deployment by logging into SALT and navigating the step-by-step deployment process. Initially, content is uploaded in a private, developmental state and subjected to a thorough review process involving automated deployment tests and code reviews by SALT administrators. This ensures that only high-quality, fully vetted content reaches production status. Following the approval, the content is made live, and the developmental version is archived, preserving the integrity and history of updates through the version-controlled repository.

Adjustments to specific content settings, such as access levels, URL customization, and scheduling of content rendering or emailing, are collaboratively managed by the publisher and SALT administrators. This meticulous process ensures that SALT continuously offers accurate, up-to-date situational awareness through a scalable and transparent system.

To deliberately uphold data privacy, SALT was designed not to store protected health information or other sensitive data within its system. Instead, it accesses necessary data sources only at the precise moment a report is generated. Processing sensitive or private data in real-time prevents the need to retain it, minimizing the risk of unauthorized access. By performing “just in time” automated data retrieval and analysis, SALT prevents human interaction with raw sensitive data. The tool exclusively shares derived analytical content, enabling informed decision-making with shared data while maintaining strict data confidentiality and compliance with privacy regulations.

Evaluation Focus Groups

The second round of focus groups aimed to evaluate the SALT prototype and included 2 sessions held in June 2024. Recruitment followed the same procedures as the first round. The interview guide included screenshots to demonstrate features of SALT and questions on how the current tool could inform decisions, potential improvements in visualization or explanation, and additions to support better decision-making (Multimedia Appendix 1). Sessions were conducted, and data were analyzed using the methods described earlier.

Insights generated from these focus groups were used to summarize the usability and effectiveness of SALT. We also compiled and summarized general suggestions for future development of SALT.

Ethical Considerations

This work did not involve engagement with human subjects research by the academic investigators as defined under US federal regulations (45 CFR 46). Participant recruitment, focus group facilitation, and qualitative data analysis were conducted by a professional research firm. The academic investigators did not participate in recruitment, did not interact with participants in the context of the focus groups, and did not have access to identifiable or raw qualitative data, receiving only anonymized, aggregated thematic findings. Although the study population comprised professionals with whom the investigators may have had prior or concurrent professional interactions unrelated to this study, no investigator-participant interaction occurred for the purposes of data collection or analysis. Accordingly, institutional review board approval for the academic investigators was not required.

Results

Design Focus Groups

We engaged approximately 20 individuals from 9 Greater Cincinnati public health and EMAs (Cincinnati Health Department, Hamilton County Public Health, Butler County Public Health, Clermont County Public Health, Highland County Public Health, Hamilton County Emergency Management, Butler County Emergency Management, Clinton County Emergency Management, and Warren County Emergency Management) in guided question-and-answer discussions to guide the development of SALT. Those participating in these initial design focus groups identified several themes related to data elements, sources, and information that they used or would find useful for situational awareness and learning in future public health events (Table 1). The groups emphasized that multiple individuals and organizations received various data forms, necessitating the need for multiple dashboards tailored to specific audiences and uses. Effective data management required integrating data from diverse sources, such as health information

XSL•FO
RenderX

exchanges, state vaccine registries, and local health care facilities, which highlighted the need for preestablished legal and infrastructure frameworks to facilitate timely data sharing. This need became especially evident as many of the emergency data connections established during the pandemic dissolved once the emergency declarations ended, leading to delays in responsive actions due to the time required to secure new data sources and legal agreements.

Table . Design requirements for the situational awareness learning tool (SALT), distilled from feedback generated during design focus groups of leaders from various agencies and organizations involved in the pandemic response in counties contained in the Ohio Hospital Association's COVID-19 Zone 6, including city and county public health departments, hospitals, and emergency management agencies.

Requirement	Description
Data compilation	Data should be compiled from multiple sources across the health care spectrum and beyond and submitted to a centralized host where data can be matched, geocoded, aggregated, and analyzed.
Legal framework	The legal framework for sharing data through SALT should reflect use cases and appropriate security and enable the wider use of sensitive, proprietary, or nonopen data.
Non–health care data	Data from regional non–health care organizations, such as schools and wastewater surveillance, are needed for informed decision-making.
Tools and dashboards	Tools and dashboards should be tailored to specific roles and audiences, with at least 2 levels: high level for community- or region-based action and detail view for person-, neighborhood-, or census tract–based intervention.
Update frequency	Updates should be made at least weekly for high-level reporting and more frequently for detailed reporting.
Surveillance measures	Surveillance measures to include and trend over time should cover infection rates, testing counts and rates, hospitalization counts and rates, mortality counts and rates, long-term care facility infection counts and rates, students in quarantine counts and rates, viral shed from wastewater surveillance rates, hospital staffing counts, and vaccination rates.
Disaggregation	The ability to disaggregate regional measures by demographics and social determinants of health would be beneficial for targeting resources and understanding potentially inequitable outcome distributions.
Accessibility	Electronic media and dashboards should be 508 compliant and accessible to users with disabilities.

The discussions also addressed the application of data in decision-making, underscoring the importance of having clear, timely, and geographically relevant data to guide the allocation of scarce resources such as personal protective equipment, laboratory kits, and human resources. Visualizations that included trend forecasting and clear action thresholds proved vital for planning and resource allocation, helping to define when and where to deploy resources based on set criteria. Moreover, the need for a centralized data hub became apparent, with frustrations voiced over the labor-intensive nature of data compilation and the inefficiency of the existing disparate systems. Participants described the ideal data tool as intuitive, tailored to specific user needs, and capable of providing real-time, actionable insights with minimal user manipulation required, ensuring rapid response capabilities in future emergencies.

One theme underscored that the perceived relevance and usefulness of data and information depend on the audience and the purpose of the tool or dashboard that displays them. Another theme was that access to data from multiple sources is essential, but it requires a legal and infrastructure framework that supports data sharing and integration before a crisis occurs. The participants also emphasized that data and information should inform decision-making and resource allocation based on identified needs, trends, forecasts, and thresholds. Furthermore, participants highlighted the importance of data visualizations that are clear, intuitive, and actionable, with definitions and annotations to avoid misinterpretation. Additionally, they stressed the need for data sharing among different groups (eg, schools and hospitals) for coordinated and effective response. Moreover, they expressed the challenge of compiling and interpreting data, which can be resource intensive and time consuming, and suggested the need for a singular, trustworthy data feed that reduces the burden and complexity of data analysis. Finally, they pointed out the gap between data received and action to take, which was often lagging and inefficient, even as events on the ground called for timely and near-real-time data.

Focus group participants identified many sources of data used during the COVID-19 pandemic. All group sessions independently identified that additional data sources may have been used; however, some sources could not immediately be recalled during the focus group sessions as they may have been used for short-term analysis or transitioned to alternate sources. Among the data sources cited were health information exchanges for COVID-19 laboratory testing, admission, and discharge data; publicly available data sources such as the Centers for Disease Control and Prevention, World Health Organization,

XSL•FO
RenderX

and the New York Times; water treatment facilities and sewage data with COVID-19 viral testing; and public school reporting with quarantine and isolation status. Data were also drawn from personal protective equipment inventories, including hospitals, health care providers, community organizations, and EMAs; long-term care facilities; skilled nursing facilities; Ohio Department of Health; Ohio Hospital Association; Greater Dayton Area Hospital Association; Ohio Disease Reporting System; Empower Data, which identifies residents dependent or homebound with medical devices; and vaccination registries.

Platform Development

To optimize the flexibility and real-time capabilities of our system, we chose to use a commercial platform designed for code-based analytical content sharing and management. This platform supports the development of interactive, data-driven solutions using R (R Foundation) or Python (Python Software Foundation), integrating these applications within a version-controlled environment (Git) to facilitate continuous integration and testing.

SALT has been effectively implemented within a VPC, and the methodology for content publication has demonstrated its effectiveness via its use in regular briefings in the state response structure, including the regional Hospital Steering Committee and the regional Multi-Agency Coalition (health commissioners, hospital leaders, and state EMAs). It is also used in regular meetings of state health-related professional societies and in local and regional meetings of informal physician and nursing groups, including the Skilled Nursing Facility and Hospital Discharge Group. Additionally, SALT has replaced the manual, weekly process of posting a readable document to a website with an automated process.

The SALT landing page serves to introduce the tool and serve as a portal for future collaborations between potential partners and the SALT team. Upon signing in, publishers can manage their version-controlled content and adjust access settings, while authorized viewers can access permitted content.

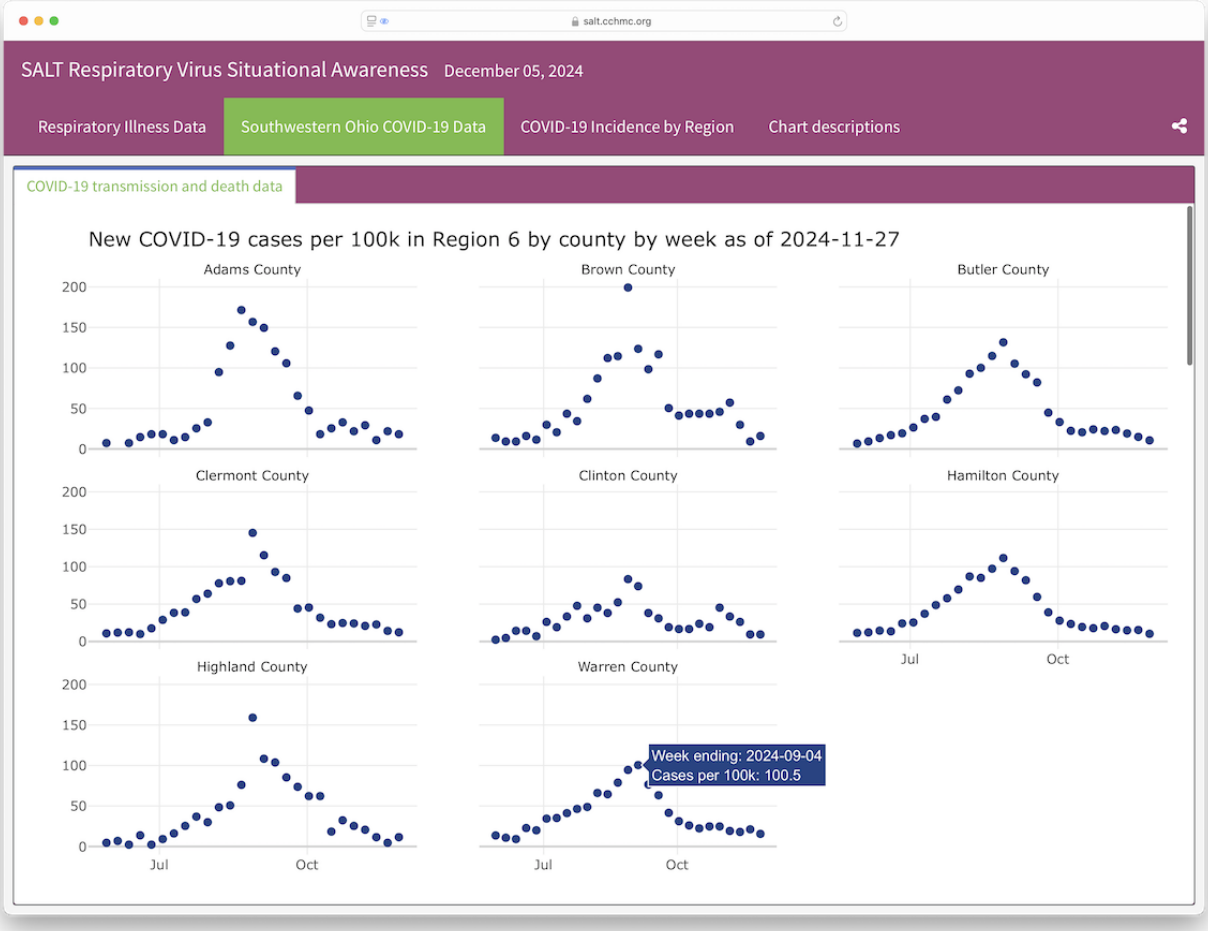
For user management, we developed an internal dashboard to monitor usage metrics around specific items and publisher activities. This dashboard provides weekly updates on user accounts, content views, and resource use. The data collected are used in a charge-back system designed to offset computing and licensing costs effectively.

Respiratory Illness Situational Awareness

A notable feature available to the public is the Greater Cincinnati COVID-19 dashboard (Figure 1). Previously a collection of static files and images, this publicly available dashboard is now hosted online [17], featuring interactive visuals and a contemporary tabbed layout. It has become an instrumental resource for regional hospital executives and senior public health administrators, aiding in effective resource management and enhancing care delivery across the region.

XSL•FO
RenderX

Figure 1. A screen capture of the publicly available SALT respiratory situational awareness dashboard [17]. The currently displayed tab highlights data specific to COVID-19 in southwestern Ohio. SALT: situational awareness and learning tool.



The regional Hospital Steering Committee, founded early in the COVID-19 pandemic, continues to meet virtually periodically and uses SALT for situational awareness of community respiratory disease. The Hospital Steering Committee dashboard is offered as a distinct and separate area on SALT because it is purpose designed to provide this user group with specific information according to a standardized narrative structure. The committee selected this narrative structure for briefings early in the COVID-19 pandemic in which a sequence of data was presented first at the national level, followed by state-level data, followed by regional- and local-level data. In this way, users gain both a sense of awareness of disease activity locally and of threats in other places that may materialize in the local region of interest. Authorization restricts viewership of data that cannot, due to sensitivity or regulatory reasons, be shared with non–Steering Committee SALT users. This area is frequently used to present brief updates in virtual meetings; committee members can also check the dashboard for updates between committee meetings.

Evaluation Focus Groups

The aggregated discussion themes and key elements from the focus groups highlight the perceived utility and areas for improvement of the SALT (Table 2). Participants affirmed SALT's potential effectiveness in managing new respiratory illness pandemics, noting its role as a reliable, consistent, and timely information source that could enhance decision-making and strategic planning, particularly with its ability to track infection rates and hospital occupancy rates, and identify vulnerable populations. Key feedback emphasized the need for more efficient resource allocation, with suggestions to include more specific analytics, such as medical equipment data, to improve actionable outcomes. Moreover, the focus groups suggested enhancements to the tool's usability through better visualization clarity, narrative explanations, and the incorporation of additional data elements, such as the availability of personal protective equipment and more granular geographic details. Recommendations also covered the integration of predictive analytics and sector-specific data to broaden SALT's applicability and enhance its utility in public health management.

XSL•FO
RenderX

Table . Improvements for the situational awareness learning tool, distilled from feedback generated during design evaluation focus groups of leaders from various agencies and organizations involved in the pandemic response in counties contained in the Ohio Hospital Association's COVID-19 Zone 6, including city and county public health departments, hospitals, and emergency management agencies.

Improvement	Description
Clearer data labeling and improved visual design	Add more descriptive titles, axes labels, and consistent naming conventions for data elements.
Expanded data coverage	Incorporate PPE[a] and vaccine inventory data and allow for more granular geographic reporting (eg, ZIP code or census tract level).
Narrative explanations	Include short narrative descriptions to accompany key graphs and charts for nontechnical users.
Forecasting capabilities	Add short-term predictive models for health system strain, particularly for ICU[b] and emergency department usage.
Legal framework for secure data sharing	Ensure the infrastructure is in place for real-time, multisector data integration before a crisis arises.

[a]PPE: personal protective equipment.

[b]ICU: intensive care unit.

Discussion

Principal Findings

Overall, we achieved our objective to build and maintain a capability that enables data sharing, rapid analysis, and situational awareness in support of the health and well-being of the regional population, both at present and in potential future pandemics or related crises. This is embodied in the cloud-based SALT, which is capable of rapidly and flexibly supporting collaboration, receiving and sharing data, analyses, and communications with and between partners at different levels of security and access.

The relevance of these core ideas was vetted with a diverse set of potential end users, who collectively expressed the need for such a system. These users described a set of system requirements that SALT designers used to select the technical solutions described in this report. For example, as described in Table 1, the user focus groups found that it is useful if data can be compiled from multiple sources; dashboards and analytic products can be tailored to specific audiences in accordance with appropriate levels of security; and updates to products can be made when data change. SALT supports these and other user requirements.

Situational awareness using appropriate, multisource data is a critical component in managing public health emergencies such as pandemics [13]. It involves the perception of environmental elements, comprehension of their meaning, and projection of their status in the near future [18]. Experience in the COVID-19 pandemic illustrated that situational awareness is essential for effective, timely public health decision-making and responses [2,8,9,19].

Moreover, SALT uses open-source scripting languages (R and Python), so that scripts generating specific analytic products can easily be shared. Over time, SALT will accumulate a library of scripts that will reduce future product development time. Similarly, user-requested revisions to analytic products can easily be accommodated at the script level, which is not always possible using commercial, proprietary visualization software. Centralized deployment of analytic content using script-based sources and processes also facilitates the monitoring of the SALT to collect and analyze usage data to continuously improve user experiences.

The ability to segment access to data according to SALT user credentials allows data to be shared easily and rapidly with those who have a validated need and permission. Nevertheless, data use agreements are still needed to access and share protected data. Given recent instabilities in the federal public health and demographic data ecosystems [20,21], it is critical for jurisdictions to begin establishing data sharing agreements across multiple health and social sector actors so that tools such as SALT can be used for situational awareness.

While SALT is still in the early stages of use and engagement, a successful launch, feedback gathering, and content deployment have been achieved. These benchmarks comprise the foundation that SALT will continue to build upon. The technical architecture and systems developed here have proven to be capable of accomplishing the objective of SALT itself, that being an effective and efficient situational awareness and learning system presenting near-real-time analytic content available to public health professionals.

A potential limitation of our broader contribution is that SALT was developed for Southwest Ohio, although the challenges it addresses are not unique to our region. The COVID-19 pandemic revealed limitations in how public health data are collected, integrated, and shared across the nation [9]. For example, local communities faced, and continue to face, similar barriers in accessing timely, multisource data; managing legal and privacy frameworks; and tailoring dashboards to different audiences. Our work demonstrates how we overcame such barriers. While the SALT pilot work occurred in Ohio, these insights are widely applicable and provide a roadmap for other jurisdictions and organizations to strengthen situational awareness and preparedness for pandemics and complex health challenges.

XSL•FO
RenderX

Conclusions

While an upcoming period of expansion and scaling will continue to evaluate the effectiveness of the tool, the systems and architecture are established and ready for widespread adoption, implementation, and success. Beyond the application described in this report related to pandemics, SALT is applicable to other health-related and health care–related domains. Learning health networks, which collect and transform multisource data into situational awareness for a range of users, including clinicians, patients, and caregivers, are but one example. A range of community health issues require comprehensive situational awareness for decision-making, such as programs improving child and maternal health equity and projects aimed at reducing asthma-related hospitalizations. It is our hope that the work described herein can and will be applied to the spectrum of health-related applications requiring shared situational awareness.

Acknowledgments

The authors would like to thank the Hamilton County Public Health and Cincinnati Health Department for their support and collaboration, which was instrumental in the success of this study.

Funding

This project is supported in whole by federal award 21.027 awarded to Hamilton County, Ohio, by the US Department of the Treasury.

Data Availability

Data on user focus design and evaluation groups are presented in the main manuscript.

Authors' Contributions

CB contributed to methodology, software, writing the original draft, and reviewing and editing. CH contributed to reviewing and editing. TM contributed to conceptualization, methodology, and resources. PK contributed to methodology, software, and reviewing and editing. AV contributed to software. AB contributed to conceptualization and methodology. DH contributed to conceptualization, funding acquisition, methodology, software, writing the original draft, and reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions used in design and evaluation focus groups.
[DOCX File, 22 KB - ojphi_v18i1e77379_app1.docx]

References

1. Lurio J, Morrison FP, Pichardo M, et al. Using electronic health record alerts to provide public health situational awareness to clinicians. J Am Med Inform Assoc 2010;17(2):217-219. [doi: 10.1136/jamia.2009.000539] [Medline: 20190067]
2. Dixit RA, Hurst S, Adams KT, et al. Rapid development of visualization dashboards to enhance situation awareness of COVID-19 telehealth initiatives at a multihospital healthcare system. J Am Med Inform Assoc 2020 Jul 1;27(9):1456-1461. [doi: 10.1093/jamia/ocaa161] [Medline: 32618999]
3. Galaitsi SE, Cegan JC, Volk K, Joyner M, Trump BD, Linkov I. The challenges of data usage for the United States' COVID-19 response. Int J Inf Manage 2021 Aug;59:102352. [doi: 10.1016/j.ijinfomgt.2021.102352] [Medline: 33824545]
4. Dong E, Ratcliff J, Goyea TD, et al. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. Lancet Infect Dis 2022 Dec;22(12):e370-e376. [doi: 10.1016/S1473-3099(22)00434-0] [Medline: 36057267]
5. Meehan AA, Flemming SS, Lucas S, et al. Data equity as a building block for health equity: improving surveillance data for people with disabilities, with substance use disorder, or experiencing homelessness, United States. Public Health Rep 2024;139(1_suppl):62S-70S. [doi: 10.1177/00333549241245624] [Medline: 38779994]
6. About public health data interoperability. CDC. 2024. URL: https://www.cdc.gov/data-interoperability/php/about/index.html [accessed 2024-07-26]
7. Cui Y, Eccles KM, Kwok RK, Joubert BR, Messier KP, Balshaw DM. Integrating multiscale geospatial environmental data into large population health studies: challenges and opportunities. Toxics 2022 Jul 20;10(7):403. [doi: 10.3390/toxics10070403] [Medline: 35878308]
8. Beck AF, Mattingly T, Brokamp C, et al. The test and protect program: a data-driven, community-engaged approach to COVID-19 testing site localization. J Public Health Manag Pract 2025;31(1):61-64. [doi: 10.1097/PHH.0000000000001993] [Medline: 39321404]

XSL•FO
RenderX

9. Stanzler M, Figueroa J, Beck AF, et al. Learning from an equitable, data-informed response to COVID-19: translating knowledge into future action and preparation. Learn Health Syst 2024 Jan;8(1):e10369. [doi: 10.1002/lrh2.10369] [Medline: 38249853]
10. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches: SAGE Publications; 2024.
11. Krueger RA. Focus Groups: A Practical Guide for Applied Research: SAGE Publications; 2014.
12. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
13. Hartley DM, Perencevich EN. Public health interventions for COVID-19: emerging evidence and implications for an evolving public health crisis. JAMA 2020 May 19;323(19):1908-1909. [doi: 10.1001/jama.2020.5910] [Medline: 32275299]
14. Hartley DM, Beck AF, Seid M, et al. Multi-sector situational awareness in the COVID-19 pandemic: the southwest Ohio experience. In: Berube DM, editor. Pandemic Communication and Resilience: Springer International Publishing; 2021:265-278. [doi: 10.1007/978-3-030-77344-1_16]
15. Zafar MA, Beck AF, Chirumbolo C, et al. A regional learning health system of congregate care facilities for COVID-19 response. Learn Health Syst 2024 Jul;8(3):e10407. [doi: 10.1002/lrh2.10407] [Medline: 39036529]
16. Barker P, Hartley D, Beck AF, et al. Rethinking herd immunity: managing the COVID-19 pandemic in a dynamic biological and behavioral environment. NEJM Catal 2021. [doi: 10.1056/CAT.21.0288]
17. Situational awareness and learning tool. URL: https://www.cctst.org/cctst/pandemic-dashboard [accessed 2026-04-13]
18. Endsley M. Designing for Situation Awareness: An Approach to User-Centered Design, 3rd edition: CRC Press; 2001.
19. Beck AF, Hartley DM, Kahn RS, et al. Rapid, bottom-up design of a regional learning health system in response to COVID-19. Mayo Clin Proc 2021 Apr;96(4):849-855. [doi: 10.1016/j.mayocp.2021.02.006] [Medline: 33714596]
20. Some CDC health data and webpages still offline after judge's order. CBS News. URL: https://www.cbsnews.com/news/some-cdc-health-data-webpages-still-offline-judge-order/ [accessed 2026-04-13]
21. Trump administration agrees to restore federal health data and websites in lawsuit settlement. PBS News. URL: https://www.pbs.org/newshour/politics/trump-administration-agrees-to-restore-federal-health-data-and-websites-in-lawsuit-settlement [accessed 2026-04-13]

Abbreviations

CFR: Code of Federal Regulations
EMA: emergency management agency
SALT: situational awareness and learning tool
VPC: virtual private cloud

Edited by E Mensah; submitted 12.May.2025; peer-reviewed by AA Meehan, ER Daly; revised version received 23.Jan.2026; accepted 04.Mar.2026; published 29.Apr.2026.

Please cite as:
Brokamp C, Hartlage CS, Mattingly T, Kuhnell P, Vancil A, Beck AF, Hartley D
Building Enhanced Public Health Data Systems With a Situational Awareness and Learning Tool: Focus Group Study
Online J Public Health Inform 2026;18:e77379
URL: https://ojphi.jmir.org/2026/1/e77379
doi:10.2196/77379

© Cole Brokamp, Carson S Hartlage, Tiffany Mattingly, Pierce Kuhnell, Andrew Vancil, Andrew F Beck, David Hartley. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 29.Apr.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Persona Development in Washington State: Mixed Methods Approach Using Statewide Survey Data

Jordan M Garcia[1], MSW; Rebecca A Hills[2], PhD; Debra Revere[3], PhD; Jaewon Lim[4], MS; Adam S Elder[2], PhD; Chris Baumgartner[5], BS; Bryant T Karras[5], PhD; Janet G Baseman[2], PhD

[1]School of Social Work, University of Washington, Seattle, WA, United States
[2]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, United States
[3]Department of Health Systems & Population Health, School of Public Health, University of Washington, Seattle, WA, United States
[4]Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, United States
[5]Washington State Department of Health, Tumwater, WA, United States

Corresponding Author:
Jordan M Garcia, MSW
School of Social Work
University of Washington
4101, 15th Ave NE
Seattle, WA, 98105
United States
Phone: 1 2096485843
Email: jordg@uw.edu

Abstract

Background: Personas, fictional profiles representing user segments, play an important role in human-centered design, ensuring tools are tailored to the needs of users. Although public health organizations often develop information systems to promote population health, human-centered design methods and personas are generally underused in public health informatics projects.

Objective: This study aims to present a novel, mixed methods approach to developing data-driven personas for use in public health information system design, leveraging 2 statewide surveys conducted in Washington State. The aim is to produce realistic, representative, and actionable personas that reflect the diversity of a state population and support user-centered design in public health initiatives.

Methods: Quantitative (cluster analysis) and qualitative (thematic review and quote extraction) methods were applied to 2 statewide survey datasets: (1) a statewide knowledge, attitudes, and practices survey (N=1103) which used random, address-based sampling, and (2) a subset of the knowledge, attitudes, and practices respondents (N=143), which included more targeted questions on opinions and preferences related to public health information systems. Characteristics examined included demographics, technological readiness, opinions about public health policies, and experience using online health tools.

Results: K-prototype clustering resulted in 5 clusters. These 5 clusters were studied using both quantitative and qualitative analysis of key factors of the Washington State population to build 13 personas. Each persona represents a different population demographic, varying levels of technological readiness and attitudes toward public health policies, and differing experiences with online health tools. Persona descriptions are further elucidated with a short profile and 2-3 quotes.

Conclusions: This study offers a scalable and adaptable framework for persona development in public health, demonstrating how existing datasets can be transformed into effective design tools. Through a mixed methods approach, personas that reflect the diverse needs, preferences, and behaviors of Washington State residents were created. These personas can enhance the design, development, and evaluation of public health information systems by centering on user experience. Persona development and the methods described here can be used in future public health informatics projects to assist in formative research, guide design and development, inform usability testing, and shape communication strategies. By bridging the gap between large-scale data and user-centered design, this approach provides a practical model for making public health technologies more aligned with community needs.

(Online J Public Health Inform 2026;18:e75422) doi:10.2196/75422

XSL•FO
RenderX

KEYWORDS

personas; user-centered design; public health informatics; information system design; cluster analysis; qualitative data

Introduction

Public health informatics projects face the challenge of designing tools that resonate with diverse populations while being applied universally [1,2]. Due to time pressures and limited resources of public health work, it is often difficult to gather sufficient user input to gain a nuanced understanding of the needs of the population being served. Further, surveys and population summaries may not adequately capture the complexity and variability in public attitudes and behaviors [3,4]. User-centered approaches can help bridge this gap by focusing on the specific needs of individuals [5]. One such user-centered design (UCD) tool, personas, can assist in representing individual needs [6]. This study focuses on the creation of detailed personas using a rich dataset, offering a novel approach to public health informatics design work that balances empirical data with granular insights.

Personas are fictional representations of individual users and may be best known for their use in marketing; however, personas are also used by informaticians, designers, and developers to better understand the needs, behaviors, and preferences of potential users of their products, systems, or services [7]. Generally, personas may include demographic information and behavioral patterns derived from user research. As part of a UCD approach, personas foster empathy with users, generating valuable insights for product design, development, and strategy [6]. Essentially, personas can provide a more concrete and actionable representation of both unstructured and complex user data, making it easier to interpret.

Although personas are widely used in information systems design, their application in health IT and public health informatics remains limited [8-11]. Public health efforts, in particular, could benefit from the use of personas, especially in contexts where resources to support direct user engagement and data collection are constrained. In public health, personas may be useful in various processes: the design and evaluation of products, recruiting for usability testing activities [12], and facilitating communication across stakeholder groups during development and decision-making processes [13]. In particular, personas can be an innovative strategy for public health teams by offering relatable, data-driven representations of target populations. This approach can enhance the reach and effectiveness of public health initiatives by keeping user needs at the forefront of design and implementation. Additionally, personas can support stronger communication and marketing efforts, making public health campaigns more engaging and accessible to diverse audiences. By integrating personas, teams may create more responsive, user-centered solutions that can improve public health outcomes.

Traditional methods of persona development often rely on small sample sizes or limited information, leading to less accurate and less representative user profiles [14]. Additionally, there is no standardized methodology, resulting in inconsistencies and differing priorities in the development process [15]. Some approaches prioritize rich qualitative data, obtaining the data from a very limited sample of the population. While these processes are designed to capture detailed individual experiences and perspectives, they may miss insights from unsampled portions of the population or key overarching trends [16]. In contrast, development processes focused on quantitative methods may use larger pools of user data, but generalizations may occur, and subtleties can be lost [17,18]. Mixed methods approaches seek to find a balance between these 2 [6]. This study emphasizes the importance of a comprehensive approach that integrates both robust quantitative metrics and detailed qualitative insights to create a more extensive and nuanced understanding of user experiences and needs.

The objective of this study is to develop a set of personas for use in public health informatics initiatives using a mixed methods approach. By synthesizing quantitative and qualitative data from representative datasets, this study aims to create realistic and representative personas that can inform the design and implementation of public health technologies. These personas are intended to support various project stages, including formative research, usability testing, and communication strategies, ensuring that public health technologies are tailored to meet the specific needs, preferences, and barriers faced by diverse population segments. This study demonstrates the efficacy of integrating quantitative and qualitative data in persona development, leveraging an approach that combines statistical rigor with contextual insights. The resulting approach provides a versatile template adaptable for other public health initiatives.

Methods

Context

WA Verify, developed by the Washington State Department of Health, is a smartphone-based tool launched in November 2021 that provides Washington residents with digital access to their COVID-19 vaccine records. To evaluate it, a statewide survey assessed knowledge, attitudes, and practices regarding public health technologies. Findings highlighted both high rates of internet and smartphone access and concerns about privacy and record reliability (Multimedia Appendix 1).

As pandemic restrictions eased, a second-phase survey was conducted in April 2023 to explore potential expanded uses for WA Verify and its underlying technology. This survey examined preferences and concerns related to privacy and security. Findings revealed that participants generally trusted public health tools but worried about the security of their personal health data. Nevertheless, many participants acknowledged the societal benefits of these tools, ultimately weighing the advantages over perceived risks.

The 2 rich datasets described above offered a unique opportunity to understand the Washington State population as a whole and capture some of the specific concerns and opinions of Washington residents in their own words. In public health, time

XSL•FO
RenderX

and resource limitations may hinder optimal user input during the planning and design of informatics projects. To address this gap, personas emerged as 1 way to leverage the available data in order to represent users and succinctly communicate their needs and preferences. A persona is a fictitious individual created to represent a user group, providing designers with personal stories that offer insights into motivations and behaviors. While traditional UCD efforts often rely on small sample sizes and qualitative data, the 2 evaluation survey datasets with rich representative data, as well as focused qualitative data on informatics topics, presented an opportunity to create a set of data-derived personas. The goal was to design personas that exhibit the following attributes:

1. Simple and relatable: provide a collection of distinct individuals who can be understood at an intuitive and personal level to help with product design.
2. Accurate: as much as possible, accurately represent the relationship between demographic characteristics and technology-related tendencies.
3. Diverse: represent a range of characteristics. While not capturing every individual, the set of personas should encapsulate diverse attributes, encouraging consideration of various population segments.
4. Actionable: provide insights that can inform decisions and strategies in public health informatics projects.
5. Comprehensive: elucidate motivations, intentions, and needs, not just demographics.

In the development process, these persona attributes may be at odds with one another. For example, the goals of relatability and actionability may conflict with complete statistical accuracy. This approach seeks to find a balance among these goals in order to create a data-driven and useful set of personas.

Data Collection

The WA Verify evaluation project conducted 1 statewide survey and later a second-phase survey; data from these 2 surveys were used for persona development. Survey methods are summarized below and are described in detail elsewhere [19].

The initial statewide survey was mailed to a random sample of 5000 Washington households between September 2022 and September 2023, with 1491 responses (32% response rate). It assessed technology use, digital literacy, internet access, experiences with COVID-19 verification, and attitudes toward public health tools.

The second-phase survey targeted a smaller sample but aimed to gather more specific information about barriers to technology use in the public health context. The survey focused on digital literacy, internet safety perceptions, digital privacy and security, confidence in public health initiatives, and opinions on hypothetical public health tools, including some space for free responses from participants to elaborate on their views. The complete second-phase survey instrument is included in Multimedia Appendix 2. The survey sample consisted of 321 respondents from the statewide survey who expressed interest in future studies, approximately 20% of all initial respondents. Data collection took place in April 2023, yielding responses from 143 individuals (44.5% response rate).

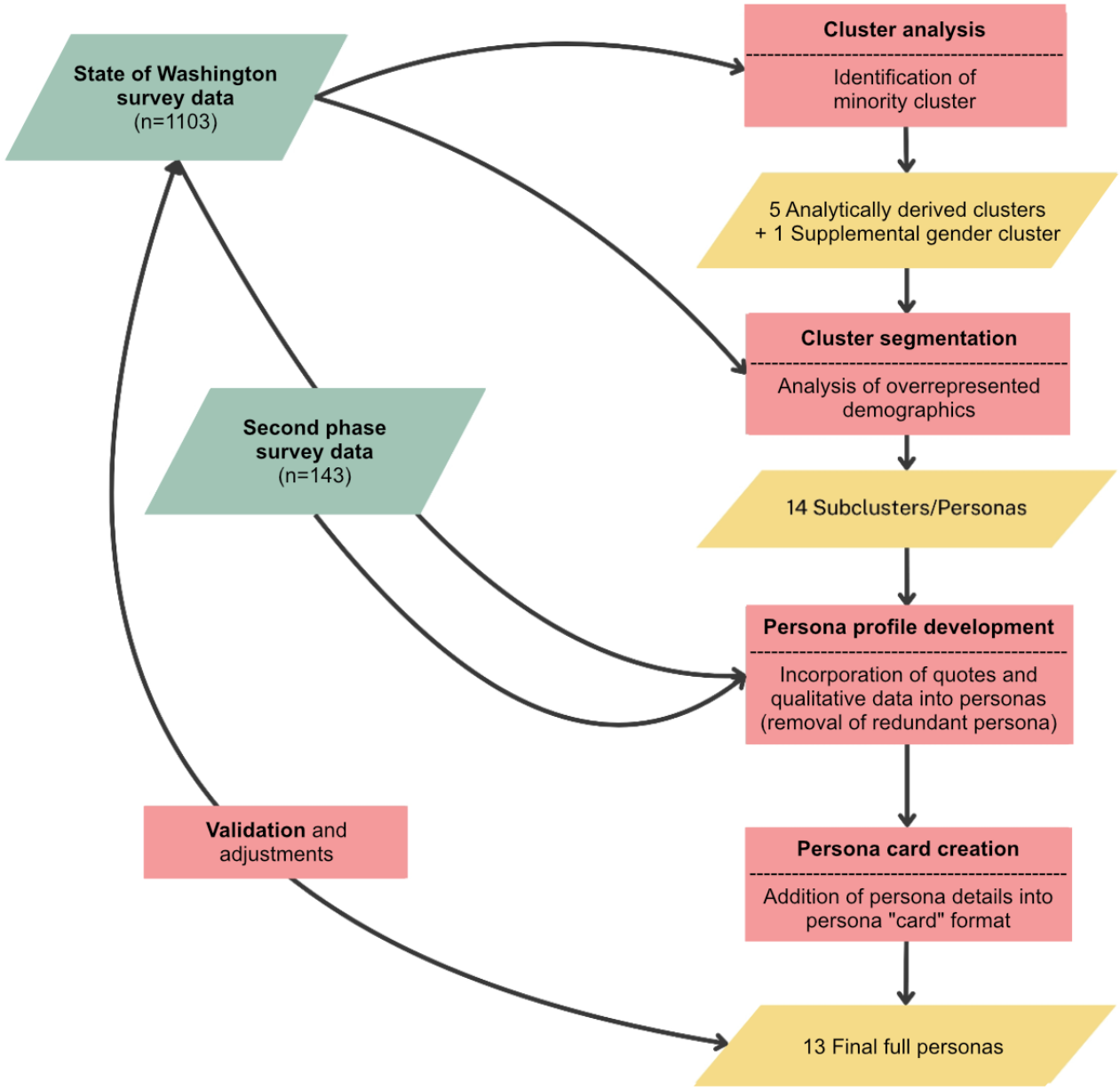
Data Analysis

Personas were developed from the 2 survey datasets using a mixed methods approach, combining quantitative and qualitative analyses. While cluster analysis provided a statistical foundation, ensuring that the final product was aligned with the survey data, this method alone did not allow for comprehensive personas. The mixed methods approach allowed for refinement of the initial clusters into distinct groups with specific demographic characteristics, aiming to capture the specific needs of diverse user groups. The process involved 5 main steps: cluster analysis, cluster segmentation, persona profile development, persona card creation, and validation. Figure 1 provides a graphical representation of this process, and each step is summarized below.

1. Cluster analysis: cluster analysis identified 5 primary clusters based on 14 key variables, including demographic, technological, and policy opinion characteristics from the State of Washington survey data. An additional cluster was created to represent nonbinary and transgender individuals who did not have sufficient representation to be included in the primary cluster analysis.
2. Cluster segmentation: quantitative methods expanded the 5 primary clusters into 14 subclusters or cluster segments. Joint distributions of age, sex, and race and ethnicity were examined to create distinct cluster segments within each cluster. Outcome variables were then summarized and assigned to these segments, creating personas. Race and ethnicity were combined into a single variable, which categorized individuals into one of five groups. All individuals who indicated they were Hispanic were categorized as “Hispanic.” Among those who did not indicate they were Hispanic, individuals who indicated they were only White, only Black, or only Asian were categorized as “White,” “Black,” and “Asian,” respectively. All other individuals (including individuals who selected multiple races) were categorized as “Other.”
3. Persona profile development: qualitative data integration enriched personas with quotes and themes extracted from both survey datasets. This process added depth and nuance to personas with additional characteristics and narrative descriptions. Redundant personas were consolidated, resulting in a final set of 13 personas.
4. Persona card creation: after defining the personas, persona cards were developed to visually represent each persona’s characteristics, needs, and motivations.
5. Validation: a validation process ensured that the personas accurately reflected the demographic and attitudinal diversity observed in the Washington datasets. Team review and quantitative comparison methods were used to iteratively refine and validate each persona.

XSL•FO
RenderX

Figure 1. Steps of the persona development process.



Quantitative Methods

Cluster Analysis

To identify distinct groups within the dataset, K-prototype clustering with Gower's distances [20] was used. K-prototype clustering is an extension of K-means that is suitable for datasets with numerical and categorical variables [21,22]. A distance function was defined to create clusters based on variables of interest, producing a "distance" value indicating similarity between respondents, with smaller values denoting greater similarity and larger values less similarity. The 14 variables used in this analysis can be seen in Table 1.

XSL•FO
RenderX

Table 1. Cluster analysis variables.

Variables	Description	Cluster analysis categories
Demographics		
Age (years)	What is your age?	18-39; 40-59; 60+
Sex	Are you _____?	Female; male; transgender; nonbinary and nonconforming; Prefer not to respond
Race and ethnicity	How would you describe your race and ethnicity?	Non-Hispanic White (White); Non-Hispanic Black (Black); Hispanic any race (Hispanic); Non-Hispanic Asian (Asian); Non-Hispanic of another race or ethnicity (Other)
Geographic region	Federal Information Processing Standard Code	Eastern Washington; Western Washington
English only at home	Do you speak a language other than English at home?	Only English; language other than English
Parental status	Are you a parent or guardian to a child under 18 years old?	Parent or guardian to <18-year-old; Not a parent or guardian to <18-year-old
Education	What is your highest level of education?	No high school degree; high school graduate; Some college or 2-year degree; 4-year degree or more
Technological and nondemographic variables		
Technological readiness	Calculated variable: overall, how confident, if at all, do you feel using computers, smartphones, or other electronic devices to do the things you need to do online? How true is the following statement? “When I get a new electronic device, or need to do a new task on it, I usually need someone else to set it up, show me how to use it, or help me.”	Higher technological readiness; lower technological readiness
eHealth tool usage	Have you ever used any electronic health tools?	Used eHealth tool; did not use eHealth tool
Method of learning of WA Verify	How did you first hear about WA Verify?	Health care; never heard; news; other; referral
WA Verify usage	Calculated Variable: Do you use WA Verify? How willing would you be to use a portable electronic COVID-19 vaccine record?	Yes; willing; not willing
COVID policy opinion	How do you feel about policies that require proof of vaccination or a negative COVID-19 test result to enter spaces that are high risk of COVID-19 spread?	Support; oppose
Health portal usage	Have you ever used any electronic health tools?: An online patient medical record or health portal (eg, MyChart).	Yes; no
WA Notify usage	Have you ever used any electronic health tools?: WA Notify (WA's COVID-19 Exposure Notification Tool).	Yes; no
Tracking apps usage (tracking)	Have you ever used any electronic health tools?: A tool for keeping track of my activities like diet and exercise (eg, Fitbit, Strava, or MyFitnessPal)	Yes; no

Since the personas were specifically developed in the context of informing the design and implementation of digital public health tools, it was crucial to differentiate between groups based on technology-related factors. To ensure that the cluster reflected a range in technology adoption, technology-related variables were weighted more heavily (5:1 relative to demographic variables). Exploratory testing with alternative weights (eg, 3 and 10) yielded similar results, with 5:1 providing a balanced emphasis on them.

Construction of the clusters was carried out by maximizing the 50-fold cross-validated silhouette score [23]. The silhouette score, ranging from –1 to 1, measures how similar each data point is to its own cluster compared to other clusters, with higher scores indicating better-defined clusters. This approach enhances the reliability and accuracy of the results by allowing for the assessment of clustering performance on data not used in its construction, reducing overfitting risk, improving cluster stability, and enhancing generalizability across different data subsets. The optimal number of clusters was determined to be 5, balancing granularity and meaningful segmentation within the dataset.

Addition of a Supplemental Cluster

Special methods were used to ensure the inclusion of demographic groups with limited presence in the sample, particularly nonbinary and transgender individuals. Due to the small number of survey respondents identifying as nonbinary

XSL•FO
RenderX

or transgender, the primary cluster analysis focused exclusively on respondents identifying as male or female. To address this limitation and ensure representation of the gender diversity present in Washington State, an additional, supplemental cluster (supplemental gender cluster) was created, including only respondents who identified as nonbinary or transgender. The cluster's summary characteristics were defined using an aggregate of the characteristics from the 8 respondents who comprised this group.

Cluster Segmentation

Because the initial clusters still contained substantial diversity, they were further segmented by age, race and ethnicity, and sex, which are factors that are known to influence behaviors and attitudes toward health tools [24-29]. Overrepresented demographics within clusters were split into additional subgroups, yielding 13 personas.

An example of the segmentation process can be seen in Table 2, which compares the distribution of the entire sample to that of Cluster 1. Cluster 1 had a higher proportion of White, male, 40- to 59-year-olds as well as Asian, female, 40-59-years prompting division into 2 subclusters, each representing one of these identified groups (White, male, 40-59 years and Asian, female, 40-59 years).

Table 2. Three-way table of age, sex, and race and ethnicity among the total survey sample (n=1103) and Cluster 1 (n=272) with observed/expected for cluster categories.

Age and sex	All		White		Black		Hispanic		Asian		Other	
	Survey, n	Cluster 1 (O/E), n (%)	Survey, n	Cluster 1 (O/E), n (%)	Survey, n	Cluster 1 (O/E), n (%)	Survey, n	Cluster 1 (O/E), n (%)	Survey, n	Cluster 1 (O/E), n (%)	Survey, n	Cluster 1 (O/E), n (%)
18-39 years												
Female	182	46 (102)	106	27 (103)	8	2 (101)	25	5 (81)	18	6 (97)	25	6 (97)
Male	110	22 (81)	67	12 (73)	2	0 (0)	6	2 (135)	19	5 (107)	16	3 (76)
40-59 years												
Female	208	74 (144)	155	53 (139)	5	0 (0)	16	4 (101)	19	11 (235)[a]	13	6 (187)
Male	136	59 (176)	106	49 (187)[a]	5	1 (81)	2	1 (203)	11	4 (147)	12	4 (135)
60+ years												
Female	259	43 (67)	225	38 (68)	1	0 (0)	7	0 (0)	4	1 (101)	22	4 (74)
Male	208	28 (55)	173	22 (52)	7	0 (0)	1	0 (0)	11	3 (111)	16	3 (76)

[a]Results were larger than expected if clusters were assigned at random across race and age.

This segmentation process, applied to each of the 5 quantitatively derived clusters, resulted in 13 distinct personas. When combined with the Supplemental Gender Cluster, which remained a standalone persona without further subdivision, a total of 14 personas were created. Outcome variables of interest (eg, technological readiness, WA Verify usage, etc.) were identified for each persona. Most subclusters were defined by the demographic characteristics determined during cluster segmentation and the most frequently observed values of the outcome variables of interest within that subcluster.

Additionally, some personas were modified to better capture the diversity of demographics and outcome variables in the population. The distribution of variables in subclusters was compared to the overall statewide population to ensure representativeness. Since each persona can only reflect 1 value per variable, it is challenging to establish actionable, diverse, and representative personas. To address this, less common subcluster characteristics were sometimes selected over more common characteristics to improve diversity and representation. Further modifications are detailed in the Validation section, ensuring that the personas are data-driven and representative of Washington State's population.

Qualitative Methods

Analysis Goals

Qualitative analysis added depth to the personas by incorporating participant quotes and themes to capture user perspectives, needs, and experiences. Thematic analysis identified recurring themes across responses, revealing common sentiments. This process added qualitative depth and real-world context to each persona.

Persona Profile Development

Personas were then synthesized and refined to ensure realistic, data-grounded representations of individuals. Key aspects crucial for constructing comprehensive personas [30] were identified and aligned with best practices for persona development [31,32], then tailored to incorporate information specifically relevant to public health information system efforts. These aspects included pertinent demographic information, public health opinions, quotes, and drivers (including needs and motivations).

Persona development and refinement took place as an iterative process with 3 University of Washington (UW) team members reviewing personas, examining quantitative and qualitative data, and reviewing and revisiting personas. The following additions

XSL•FO
RenderX

and adjustments were made to personas during the iterative refinement process (not necessarily in this order):

1. Additional stories were built around the personas based on qualitative survey data from individuals with matching or similar characteristics to the persona.
2. Personas were given exact ages, marital statuses, and occupations;
3. Quotes and variations on quotes were assigned to personas from survey respondents with similar characteristics.
4. New quotes were created where only partial ideas existed in the qualitative data.
5. Personas were given names and pictures.
6. One persona was removed (cluster 2, persona 7) as seen in Table 3

.

Table 3. Primary persona list derived from expansion of cluster analysis groupings.

Persona number	Cluster number	Age group (y)	Sex	Race or ethnicity	Region	English only	Parent	Education	Tech readiness	WA Verify refer	WA Verify usage	Opinion COVID policies	Health portal	WA Notify usage	Tracking
1	1	40-59	Male	White	W[a]	Yes	No	4-year	High	Referral	Yes	Support	Yes	Yes	Yes
2	1	40-59	Female	Asian	W	No	Yes	4-year	High	Referral	Yes	Support	Yes	Yes	Yes
3	2	18-39	Female	White	W	Yes	No	4-year	High	Referral	Willing	Support	Yes	Yes	Yes
4	2	18-39	Male	White	W	Yes	No	4-year	High	Never heard	Willing	Support	Yes	Yes	Yes
5	2	18-39	Male	Black	W	Yes	No	4-year	High	Never heard	Willing	Support	Yes	Yes	Yes
6	2	18-39	Female	Hispanic	W	Yes	No	≤2-year	High	Never heard	Willing	Support	Yes	Yes	Yes
7	2	18-39	Male	Asian	W	Yes	No	4-year	High	Never heard	Willing	Support	Yes	Yes	Yes
8	3	60+	Female	White	W	Yes	No	≤2-year	Lower	Never heard	Not Willing	Support	Yes	No	No
9	3	60+	Male	White	W	Yes	No	≤2-year	Lower	Never heard	Willing	Support	Yes	No	No
10	4	60+	Female	White	W	Yes	No	4-year	Lower	News	Willing	Support	Yes	Yes	Yes
11	5	18-39	Male	White	W	Yes	No	4-year	High	Never heard	Willing	Oppose	Yes	Yes	Yes
12	5	40-59	Female	Black	W	Yes	No	4-year	High	Referral	Willing	Support	Yes	No	Yes
13	5	18-39	Female	Hispanic	E[b]	No	Yes	≤2-year	High	Never heard	Willing	Oppose	Yes	No	Yes
14	Supp.	18-39	Nonbinary, Transgender	White	W	Yes	No	4-year	High	Never heard	Willing	Support	Yes	Yes	Yes

[a]W: Western Washington.

[b]E: Eastern Washington.

During the refinement process, representative quotes and key motivations (drivers) were added to each persona. Quotes were selected during the qualitative analysis and matched to demographically similar personas, with minor adjustments for clarity and brevity. Drivers emerged from thematic analysis and were grounded in survey data. Additionally, as mentioned above, 1 persona was removed because it reflected too small a respondent group and overlapped considerably with others from the same cluster. In line with our refinement criteria, we kept only personas that were both supported by sufficient data and represented distinct needs and motivations. Throughout this process, the balance between sufficient detail and making realistic personas while avoiding stereotypes and oversimplification was carefully considered.

Persona Card Creation

Following the development of the demographic profiles, persona cards were created to serve as visual and textual representations of the personas. The information about each persona derived from the steps previously outlined was translated into a concise

XSL•FO
RenderX

format to facilitate their use in design and decision-making processes. First, a standard template was identified to outline the sections for demographics, public health opinions, quotes, and drivers. The persona demographic profiles were transferred into this template. Each persona card also included a representative image, which was selected to be demographically appropriate and aligned with the persona's profile. After the card drafts were established, they were reviewed by UW team members, looking for redundancies and inconsistencies. This process resulted in 13 persona cards.

Validation

Because the process by which the personas were developed was novel, and in some cases ad hoc, a validation step was used to confirm that the final persona characteristics were reflective of the observed data. Each persona was validated using the following steps:

1. A group of respondents was created for each persona where each record matched the following characteristics exactly: age group, sex, race and ethnicity, geographic region, language at home, parental status, and level of education.
2. In some cases, when this group of respondents was small or non-existent, a second, larger subset was also considered that included both exact matches and individuals that matched on all but 1 of the 7 characteristics mentioned in Step 1.
3. The distributions of technological readiness, opinions on COVID-19 policies, use of digital health portals, and use of health tracking software were tabulated and compared to the characteristics assigned to the given persona.
4. In cases where a mismatch between the persona characteristic and the respondent summary was observed, the personas team discussed and came to a collaborative decision on whether to make a change to the persona characteristic.

In total, 4 major discrepancies were identified through this process, and each prompted a discussion among the team. In 2 cases, the group determined the persona's characteristics should reflect the demographic groupings, and changes were made to reflect this. However, in 2 cases, the discrepancy was retained. For both affected personas, a thoughtful decision was made to retain groups who are generally less represented, likely have unique needs, and are not well understood with respect to habits and preferences regarding public health information systems (ie, low technological readiness and less experience with online health tools). These decisions reflect the belief that although some characteristics are less common in the generally tech-savvy population of Washington State, it is important to take less common needs and traits into consideration to design accessible tools for all Washington residents.

Ethical Considerations

In June 2022, the UW Human Subjects Division reviewed the WA Verify evaluation project (IRB ID: STUDY00015786) and determined it did not qualify as research under federal and state regulations, thus exempting it from UW Institutional Review Board review. The Washington State University/Social and Economic Sciences Research Center Institutional Review Board subsequently confirmed this exempt status. Survey participants were asked to confirm their agreement to participate in the survey before starting the survey. No secondary data was used in this analysis.

Results

Overview

Overall, 13 personas emerged through the development process, 12 derived from the quantitative clustering and expansion process, and 1 Supplemental Gender Cluster.

Clusters

Cluster analysis identified 5 clusters. Additionally, a supplemental cluster was defined to identify a segment of the survey population not considered in the analysis. Table 4 displays these 6 clusters. Additional summary tables of the sample for related questions from the survey are provided in Multimedia Appendix 3.

XSL•FO
RenderX

Table 4. Five quantitatively identified clusters and one supplemental gender cluster with summary characteristics.

Characteristics	All (N=1103)	Cluster 1 (n=272)	Cluster 2 (n=223)	Cluster 3 (n=302)	Cluster 4 (n=57)	Cluster 5 (n=249)	Supplemental (n=8)
Age (years), n (%)							
18-39	292 (26)	68 (25)	97 (43)[a]	7 (2.3)	0 (0)	120 (48)[a]	5 (63)[a]
40-59	344 (31)	133 (49)[a]	68 (30)	36 (12)	17 (30)	90 (36)	1 (13)
60+	467 (42)	71 (26)	58 (26)	259 (86)[a]	40 (70)[a]	39 (16)	2 (25)
Sex, n (%)							
Female	649 (59)	163 (60)	128 (57)	164 (54)	44 (77)[a]	150 (60)	0 (0)
Male	454 (41)	109 (40)	95 (43)	138 (46)[a]	13 (23)	99 (40)	0 (0)
Nonbinary	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	5 (63)
Transgender	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (38)
Race and ethnicity, n (%)							
White	832 (75)	201 (74)	156 (70)	251 (83)[a]	48 (84)[a]	176 (71)	5 (63)
Black	28 (2.5)	3 (1.1)	8 (3.6)[a]	5 (1.7)	1 (1.8)	11 (4.4)[a]	0 (0)
Hispanic	57 (5.2)	12 (4.4)	18 (8.1)[a]	8 (2.6)	0 (0)	19 (7.6)[a]	1 (13)
Asian	82 (7.4)	30 (11)[a]	21 (9.4)[a]	13 (4.3)	4 (7)	14 (5.6)	1 (13)
Others	104 (9.4)	26 (9.6)	20 (9)	25 (8.3)	4 (7)	29 (12)	1 (13)
Region, n (%)							
Eastern	205 (19)	37 (14)	35 (16)	69 (23)[a]	9 (16)	55 (22)[a]	1 (13)
Western	898 (81)	235 (86)[a]	188 (84)[a]	233 (77)	48 (84)[a]	194 (78)	7 (88)[a]
English, n (%)							
Only English	969 (88)	230 (85)	191 (86)	281 (93)[a]	53 (93)[a]	214 (86)	8 (100)[a]
Other language	134 (12)	42 (15)	32 (14)	21 (7)	4 (7)	35 (14)	0 (0)
Parent to child <18, n (%)							
Yes	271 (25)	83 (31)[a]	60 (27)	23 (7.6)	7 (12)	98 (39)[a]	0 (0)
No	832 (75)	189 (69)	163 (73)	279 (92)[a]	50 (88)[a]	151 (61)	8 (100)[a]
Education, n (%)							
4-year degree or more	657 (60)	207 (76)[a]	146 (65)[a]	123 (41)	38 (67)[a]	143 (57)	6 (75)
2-year or some college	290 (26)	53 (19)	58 (26)	101 (33)[a]	15 (26)	63 (25)	1 (13)[a]
High school graduate	123 (11)	11 (4)	18 (8.1)	58 (19)	4 (7)	32 (13)[a]	1 (13)[a]
Less than high school	33 (3)	1 (0.4)	1 (0.4)	20 (6.6)	0 (0)	11 (4.4)[a]	0 (0)[a]
Technological readiness, n (%)							
Higher	826 (75)	256 (94)[a]	219 (98)[a]	97 (32)	13 (23)	241 (97)[a]	8 (100)[a]
Lower	277 (25)	15 (5.9)	4 (1.8)	205 (68)[a]	44 (77)[a]	8 (3.2)	0 (0)
Method of learning of WA Verify, n (%)							
Referral	223 (20)	136 (50)[a]	26 (12)	17 (5.6)	10 (18)	34 (14)	1 (13)
News	144 (13)	53 (19)[a]	11 (4.9)	31 (10)	27 (47)[a]	22 (8.8)	0 (0)
Health care	117 (11)	62 (23)[a]	19 (8.5)	19 (6.3)	3 (5.3)	14 (5.6)	0 (0)
Others	47 (4.3)	17 (6.3)	11 (4.9)	8 (2.6)	3 (5.3)	8 (3.2)	1 (13)

XSL•FO
RenderX

Characteristics	All (N=1103)	Cluster 1 (n=272)	Cluster 2 (n=223)	Cluster 3 (n=302)	Cluster 4 (n=57)	Cluster 5 (n=249)	Supplemental (n=8)
Never heard	572 (52)	4 (1.5)	156 (70)[a]	227 (75)[a]	14 (25)	171 (69)[a]	6 (75)[a]
WA Verify usage, n (%)							
Yes	276 (25)	234 (86)[a]	5 (2.2)	21 (7)	7 (12)	9 (3.6)	1 (13)
Willing	560 (51)	17 (6.3)	193 (87)[a]	150 (50)	38 (67)[a]	162 (65)[a]	6 (75)[a]
Not willing	267 (24)	21 (7.7)	25 (11)	131 (43)	12 (21)	78 (31)[a]	1 (13)
COVID policies, n (%)							
Oppose	226 (20)	26 (9.6)	27 (12)	69 (23)[a]	8 (14)	96 (39)[a]	0 (0)
Support	877 (80)	246 (90)[a]	196 (88)[a]	233 (77)	49 (86)[a]	153 (61)	8 (100)[a]
Health portal, n (%)	949 (86)	262 (96)[a]	212 (95)[a]	199 (66)[a]	57 (100)[a]	219 (88)[a]	7 (88)[a]
WA Notify usage, n (%)	535 (49)	235 (86)[a]	223 (100)[a]	28 (9.3)[a]	49 (86)[a]	0 (0)	2 (29)[a]
Tracking, n (%)	660 (60)	209 (77)[a]	181 (81)[a]	24 (7.9)[a]	45 (79)[a]	201 (81)[a]	6 (75)[a]

[a]Results were larger than expected based on the full sample.

Persona Segments

From the 5 calculated clusters and the supplemental gender cluster, 14 personas were identified to represent distinct segments of the population based on race and ethnicity, age, and sex. By identifying overrepresentation of these demographic characteristics, personas were defined within each cluster. This expanded cluster-derived persona list is displayed in Table 3.

Persona Profiles

To enhance the initial persona profiles derived from cluster analysis and segmentation, qualitative analysis was conducted. This process incorporated demographic data and qualitative responses from both surveys to identify the needs, preferences, and experiences of each persona. This approach allowed us to humanize the personas. Tables 5 and 6 present the additional attributes developed through this qualitative analysis, providing a richer, more nuanced description of each persona.

Table 5. Final persona list with demographic information.

Persona number	Cluster number	Age (years)	Sex	Race and ethnicity	Region	English only	Parent	Education	Tech readiness
1	1	40-59	Male	White	Western	Yes	Yes	4-year	High
2	1	40-59	Female	Asian	Western	No	Yes	4-year	High
3	2	18-39	Female	White	Western	Yes	No	4-year	High
4	2	40-59	Male	White	Western	Yes	No	4-year	Lower
5	2	18-39	Male	Black	Western	Yes	No	4-year	High
6	2	18-39	Female	Hispanic	Eastern	Yes	No	≤ 2-year	High
7	3	60+	Female	White	Western	Yes	No	≤ 2-year	Lower
8	3	60+	Male	White	Western	Yes	No	≤ 2-year	Lower
9	4	60+	Female	White	Western	Yes	No	4-year	High
10	5	18-39	Male	White	Eastern	Yes	Yes	4-year	High
11	5	40-59	Female	Black	Western	Yes	Yes	4-year	High
12	5	18-39	Female	Hispanic	Western	No	No	≤ 2-year	High
13	Supp.	18-39	Nonbinary	White	Western	Yes	No	4-year	High

XSL•FO
RenderX

Table 6. Final persona list with profiles and names.

Persona number	Cluster number	Name	Age (years)	Marital status	Occupation
1	1	David	43	Divorced	Software Engineer
2	1	Yuki	40	Married	PT Assistant
3	2	Emily	32	Single	Social Services
4	2	Ryan	52	Single	Grocery Stocker
5	2	Desmond	37	Married	Banking
6	2	Maria	35	Married	Human Resources
7	3	Linda	66	Married	Retired
8	3	Richard	60	Divorced	Appliance Repair
9	4	Patricia	67	Married	Retired
10	5	Kyle	31	Married	Firefighter
11	5	Alicia	46	Re-married	Teacher
12	5	Isabel	21	Single	Premedical student
13	Supplemental	Riley	26	In a relationship	Graduate Student

Persona Cards

The final 13 personas were formatted into graphic single-page summaries or persona cards. Figure 2 depicts 1 persona card (Personas 5 from Table 5). Quotes and drivers were also added to the personas at this stage.

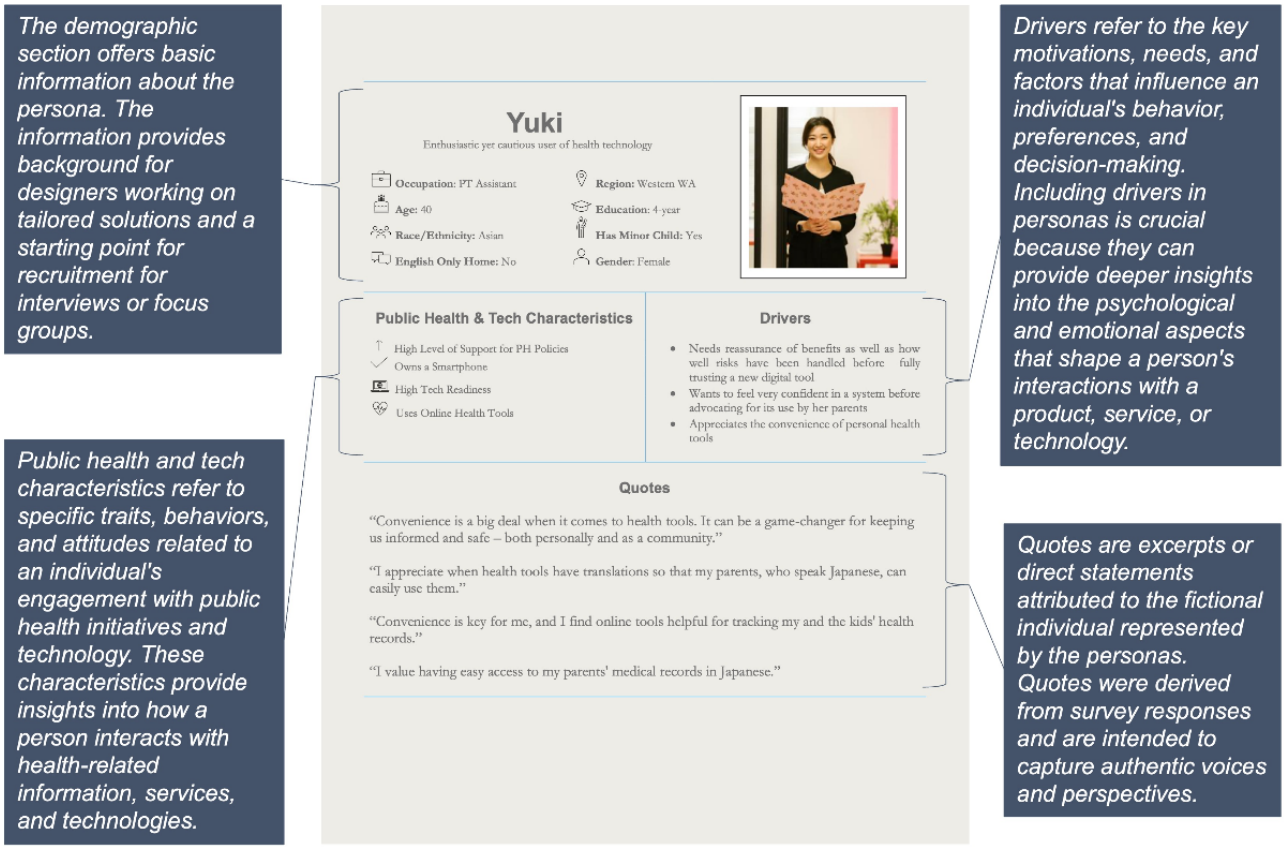
XSL•FO
RenderX

Figure 2. Persona 5: Desmond.

Aspects of the persona cards were included based on their ability to provide a practical and user-friendly tool to bridge the gap between data and actionable insights. A detailed description of the derivation and relevance of each section is provided in Figure 3.

XSL•FO
RenderX

Figure 3. Sample persona with section descriptions.



The full set of 13 persona cards can be found in Multimedia Appendix 4 and is available for download, allowing researchers and practitioners to use and adapt them for their own work.

Discussion

Principal Findings

The development of these personas leveraged a uniquely rich dataset, enabling a nuanced representation of diverse population segments. The mixed methods approach combined a rigorous data-driven process with the depth of individual stories, behaviors, and needs. This synthesis achieved a deeper understanding of the dataset and allows for quickly gleanable insights, balancing the key priorities for the persona creation, which could not be accomplished with quantitative or qualitative methods alone.

The approach focused on creating personas that were simple and relatable, accurate, diverse, actionable, and comprehensive. At times, decisions in the methodology had to be made to ensure effectiveness by finding a balance among these guiding principles. The aim was to define a collection of distinct individuals who could be intuitively understood. The personas reflect the breadth of demographic diversity in Washington State, representing a range of characteristics rather than matching exact population proportions, ensuring actionable insights for public health informatics projects.

Cluster analysis served as a starting point, ensuring the final personas were guided by the survey data. While cluster analysis alone was the most statistically accurate process, these clusters did not establish a set of personas that upheld the other key priorities. Through a mixed methods approach, distinct groups with more specific demographic characteristics were constructed. This technique helps highlight various user groups who might have more specific needs. For example, some of the personas can help understand challenges for those who have lower technological readiness that could benefit from targeted strategies regarding technology adoption and usage. While this is a minority of the population, individuals with these needs must still be considered in design, development, and communication.

This methodology aimed to mitigate potential biases from both human-driven and purely quantitative approaches to persona development. Traditional persona creation often relies heavily on individual interpretation, which can introduce unconscious human biases and assumptions. Conversely, purely data-driven approaches can amplify existing biases in the source data, which can result in further marginalizing groups underrepresented in the dataset. The mixed methods approach used in this study leverages the strengths of both while working to mitigate their weaknesses. The cluster analysis provided a foundation that reduced subjective biases in the initial grouping, while the qualitative analysis preserved a human perspective, allowing for the creation of authentic and relatable personas that addressed gaps in the quantitative results. This approach ensures statistically grounded and representative personas.

This development process considered a wide range of demographic and outcome variables, ensuring the inclusion of diverse population segments. This comprehensive approach is crucial in public health initiatives, where efforts impact a broad population in highly personal and individual ways. While

XSL•FO
RenderX

resource-intensive, this approach offers a useful framework for creating representative and realistic personas. The template may be used and adapted to meet the specific needs of various projects.

This methodology demonstrates the value of a mixed methods approach in capturing the complexity of public health user groups and provides an adaptable template for future initiatives.

Use of Personas in Public Health Informatics Projects

The personas developed through this process may serve as a tool to inform and guide public health informatics initiatives in the following ways:

1. Facilitate formative work: prior to embarking on a development project, personas can be used to understand potential users and recruit for interviews and focus groups, during which tailored information needs and user requirements data can be collected and analyzed.
2. Design and development: the design and development process can be aided by personas, providing an understanding of user preferences and behaviors. Decision-making can be aided by considering the impact of a tool or change on all target personas.
3. Usability testing: personas can be used to either simulate user interactions during usability testing or assist in the recruitment of diverse and appropriate usability testers.
4. Communications: personas can be used to facilitate internal communications (ie, development team decision-making) and external communications (eg, deciding on targets and strategies for education and information campaigns).

These personas can be used directly in practice. For example, a campaign focused on changing the health behaviors of Washingtonians could use these personas to consider how to tailor information to individuals with varying priorities and levels of digital comfort. Application of these personas can bring in considerations of real user opinions.

Implications for Public Health Initiatives

Understanding the needs and motivations behind different personas can enhance the effectiveness of public health interventions. Strategies can be tailored to specific personas, and more targeted and impactful campaigns and technologies can be implemented. A user-centered design process that is still based on comprehensive data enhances usability and ensures that interventions resonate with distinct segments of the population. This may result in higher engagement and better outcomes.

Further, as new data become available or as public health priorities shift, personas can be updated or added to reflect these changes, ensuring that the technology remains relevant and responsive to current needs. This adaptability is particularly valuable in the rapidly evolving landscape of public health, where timely and accurate information is crucial.

Moreover, the iterative and flexible nature of persona development allows for continuous refinement and adaptation. The personas developed in this study can serve as a template that other public health initiatives can modify according to their specific needs and available resources. Even without access to extensive datasets, personas can be developed using smaller datasets, targeted surveys, focus groups, or stakeholder interviews. This process and the final product can be adapted to whatever resources are available and relevant to the effort. This adaptive approach ensures that initiatives remain timely and responsive and allows for a meaningful tool across varying resource availability and project scales.

Limitations

There are several important limitations to consider regarding both the development and potential use of these personas. Most importantly, while personas are valuable tools for design and development, they are not substitutes for direct user engagement and feedback. These personas should ideally be used in conjunction with, rather than in place of, real user data and testing. They are meant to serve as practical tools when user engagement is challenging or impossible due to resource constraints, but they should not be seen as comprehensive replacements for actual user research.

Additional limitations stem from the survey data on which these personas are based. First, relying on survey data introduces sampling bias, particularly in the case of the second-phase survey data, as survey respondents may not be representative of the diverse population of Washington State. Second, survey data were collected in the context of the COVID-19 pandemic, a moment in time. Public opinions rapidly evolve, and the sentiments expressed in the surveys and subsequently used to develop the personas may only reflect that context and not be applicable to a new context. In particular, survey participation during the pandemic may have been lower among certain groups, such as lower-income populations [33], and public attitudes at the time reflected heightened concern about health and strong adherence to public health guidelines compared to non-pandemic periods [34]. These factors may limit the generalizability of the personas to future public health informatics topics. Third, the surveys focused on capturing opinions regarding WA Verify and its potential expanded use cases, so generalizing to other public health tools will depend on each tool and its context.

The methodology used presents additional limitations. Cluster analysis requires incorporating assumptions and parameter choices that may impact their interpretation, reliability, and generalization outside the context of the analysis. The development of personas may inherently oversimplify the complexity of each individual represented. In addition, bias may be introduced into the personas by including demographic and other characteristics in their development.

The addition of the supplemental gender cluster, while intended to include individuals who identify as a minority gender group, presents another limitation. Personas are not meant to be statistically representative of the population, but their value is in simplifying a complex population into a small set of fictional individuals that can be connected with. However, this simplification comes with the cost of not capturing the breadth of experiences across the population. While the representation of gender-diverse individuals in this study does not fully capture the range of identities in Washington State and oversimplifies, it offers an important point of visibility and inclusion, ensuring

XSL•FO
RenderX

these perspectives are considered within the broader set of personas.

Future Work

The overall usefulness of these personas would benefit from an iterative process of feedback with stakeholders, including developers, designers, communication teams, and any other potential users of the personas. Stakeholders who are experienced with using personas, as well as those new to using personas, would provide valuable feedback. This process can help to ensure the relevance and applicability of these tools.

In particular, 1 question to be addressed is the appropriate level of detail that is useful for these personas. In UCD, there are differing opinions on the optimal level of detail in personas. Emerging ways of thinking promote more pared-down personas where there is no image and fewer personal details, focusing instead on only the practical elements that directly affect usability [35,36]. Others advocate for more life-like personas with carefully considered details [37,38]. Some details added to a persona may make the fictitious person more tangible, but some argue that these details could be irrelevant, extraneous, or distracting, potentially even introducing unnecessary bias. In this study, we used a more complete level of detail, incorporating demographic characteristics as well as opinions related to the topic at hand. The aim was to provide enough context to support empathy, design decisions, and communication strategies when access to user interviews was minimal. Future work should explore the impact of different levels of detail in personas on the tool's effectiveness in public health informatics projects.

While evaluating the personas and their use within live public health informatics projects was beyond the scope of this study, evaluation is an important step. Future work should include pilot testing with designers, developers, and public health practitioners to assess the personas' usefulness and their impact.

Additionally, more can be done to ensure the inclusiveness of personas. This can be accomplished with further user inquiry focusing on historically marginalized groups. Persona Cards could also be iteratively reviewed with groups of people whom they presume to represent to check for validity and identify any constructed biases. Without additional resources to conduct focus groups or user research, other processes can be implemented to better represent these communities, such as a backwards user journey [30].

Conclusions

This study describes the process of developing personas from existing survey datasets. The mixed methods approach, combining quantitative cluster analysis with qualitative data integration, resulted in 13 detailed personas. These personas may be helpful in various stages of public health technology initiatives, including needs assessment, design, roll-out, and evaluation. By bridging the gap between large-scale data and actionable insights, these personas offer a powerful tool for creating more user-centered and effective public health information systems. The methodology presented here provides a flexible framework that can be adapted to meet the specific needs of various public health initiatives, potentially enhancing the use of UCD principles and thus the user-centeredness and effectiveness of public health technologies.

Funding

This work was funded through an Interagency Agreement between the Washington State Department of Health and the University of Washington School of Public Health (HED contract 25742).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Statewide survey instrument. The complete 32-question survey instrument used for statewide data collection.
[PDF File (Adobe PDF File), 462 KB - ojphi_v18i1e75422_app1.pdf]

Multimedia Appendix 2
Second phase survey instrument. The complete 53-question survey instrument used for the second phase of data collection.
[PDF File (Adobe PDF File), 70 KB - ojphi_v18i1e75422_app2.pdf]

Multimedia Appendix 3
Additional descriptives.
[DOCX File , 31 KB - ojphi_v18i1e75422_app3.docx]

Multimedia Appendix 4
Persona cards. The complete set of 13 persona cards developed through this research.
[PDF File (Adobe PDF File), 691 KB - ojphi_v18i1e75422_app4.pdf]

References

XSL•FO
RenderX

1. Aziz HA. A review of the role of public health informatics in healthcare. J Taibah Univ Med Sci 2017;12(1):78-81 [FREE Full text] [doi: 10.1016/j.jtumed.2016.08.011] [Medline: 31435217]
2. Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, Kilbourne EM. Public health informatics: improving and transforming public health in the information age. J Public Health Manag Pract 2000;6(6):67-75. [doi: 10.1097/00124784-200006060-00010] [Medline: 18019962]
3. Marsden N, Haag M. Stereotypes and politics: reflections on personas. USA: Association for Computing Machinery; 2016 Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose California USA p. 417-4031. [doi: 10.1145/2858036.2858151]
4. Salminen J, Jung S, Jansen B. Detecting demographic bias in automatically generated personas. USA: Association for Computing Machinery; 2019 Presented at: CHI EA '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow Scotland UK p. 1-6. [doi: 10.1145/3290607.3313034]
5. Vredenburg K, Mao JY, Smith PW, Carey T. A survey of user-centered design practice. USA: Association for Computing Machinery; 2002 Presented at: CHI '02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 20-25, 2002; Minneapolis Minnesota USA p. 471-478. [doi: 10.1145/503376.503460]
6. Pruitt J, Adlin T. The persona lifecycle: keeping people in mind throughout product design. Amsterdam: Elsevier; 2010.
7. Miaskiewicz T, Kozar KA. Personas and user-centered design: how can personas benefit product design processes? Design Studies 2011;32(5):417-430. [doi: 10.1016/j.destud.2011.03.003]
8. Holden RJ, Kulanthaivel A, Purkayastha S, Goggins KM, Kripalani S. Know thy eHealth user: development of biopsychosocial personas from a study of older adults with heart failure. Int J Med Inform 2017;108:158-167 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.006] [Medline: 29132622]
9. Massey PM, Chiang SC, Rose M, Murray RM, Rockett M, Togo E, et al. Development of personas to communicate narrative-based information about the HPV vaccine on twitter. Front Digit Health 2021;3:682639 [FREE Full text] [doi: 10.3389/fdgth.2021.682639] [Medline: 34713151]
10. Ten Klooster I, Wentzel J, Sieverink F, Linssen G, Wesselink R, van Gemert-Pijnen L. Personas for better targeted ehealth technologies: user-centered design approach. JMIR Hum Factors 2022;9(1):e24172 [FREE Full text] [doi: 10.2196/24172] [Medline: 35289759]
11. Turner AM, Reeder B, Ramey J. Scenarios, personas and user stories: user-centered evidence-based design representations of communicable disease investigations. J Biomed Inform 2013;46(4):575-584 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.006] [Medline: 23618996]
12. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
13. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
14. Chapman CN, Milham RP. The personas' new clothes: methodological and practical arguments against a popular method. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2006;50(5):634-636. [doi: 10.1177/154193120605000503]
15. Salminen J, Guan K, Jung S, Chowdhury S, Jansen B. A literature review of quantitative persona creation. USA: Association for Computing Machinery; 2020 Presented at: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu HI USA p. 1-14. [doi: 10.1145/3313831.3376502]
16. Jansen A, Van MM, Slegers K. Personas and behavioral theories: a case study using self-determination theory to construct overweight personas. USA: Association for Computing Machinery; 2017 Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver Colorado USA p. 2127-2136. [doi: 10.1145/3025453.3026003]
17. Hajian S, Bonchi F, Castillo C. Algorithmic bias: from discrimination discovery to fairness-aware data mining. USA: Association for Computing Machinery; 2016 Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco California USA p. 2125-2126. [doi: 10.1145/2939672.2945386]
18. Siegel DA. The mystique of numbers: belief in quantitative approaches to segmentation and persona development. USA: Association for Computing Machinery; 2010 Presented at: CHI EA '10: CHI '10 Extended Abstracts on Human Factors in Computing Systems; April 10-15, 2010; Atlanta Georgia USA p. 4721-4732. [doi: 10.1145/1753846.1754221]
19. Molino A, Revere D, Hills R, Elder A, West L, Karras B, et al. Exploring attitudes and obstacles around digital public health tools: insights from a statewide cross-sectional survey on washington's vaccine verification system. J Med Internet Res 2025;27:e66550 [FREE Full text] [doi: 10.2196/66550] [Medline: 41042596]
20. Gower JC. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 1966;53(3-4):325-338. [doi: 10.1093/biomet/53.3-4.325]
21. Ji J, Pang W, Zhou C, Han X, Wang Z. A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data. Knowledge-Based Systems 2012;30:129-135. [doi: 10.1016/j.knosys.2012.01.006]
22. Ji J, Bai T, Zhou C, Ma C, Wang Z. An improved k-prototypes clustering algorithm for mixed numeric and categorical data. Neurocomputing 2013;120:590-596. [doi: 10.1016/j.neucom.2013.04.011]

XSL•FO
RenderX

23. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53-65. [doi: 10.1016/0377-0427(87)90125-7]
24. Berens E, Vogt D, Messer M, Hurrelmann K, Schaeffer D. Health literacy among different age groups in Germany: results of a cross-sectional survey. BMC Public Health 2016;16(1):1151 [FREE Full text] [doi: 10.1186/s12889-016-3810-6] [Medline: 27829401]
25. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
26. Goswami A, Dutta S. Gender differences in technology usage—a literature review. OJBM 2016;04(01):51-59. [doi: 10.4236/ojbm.2016.41006]
27. Haluza D, Wernhart A. Does gender matter? Exploring perceptions regarding health technologies among employees and students at a medical university. Int J Med Inform 2019;130:103948. [doi: 10.1016/j.ijmedinf.2019.08.008] [Medline: 31442846]
28. Mantwill S, Monestel-Umaña S, Schulz PJ. The relationship between health literacy and health disparities: a systematic review. PLoS One 2015;10(12):e0145455 [FREE Full text] [doi: 10.1371/journal.pone.0145455] [Medline: 26698310]
29. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
30. Spillers F. How to create inclusive personas. How to create Inclusive Personas.: Frank Spillers & Experience Dynamics; 2023 Nov 24. URL: https://frankspillers.com/how-to-create-inclusive-personas/ [accessed 2024-07-12]
31. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
32. Salminen J, Guan K, Jung S, Jansen BJ. A survey of 15 years of data-driven persona development. Int J Hum Comput Interact 2021;37(18):1685-1708. [doi: 10.1080/10447318.2021.1908670]
33. Krieger N, LeBlanc M, Waterman P, Reisner S, Testa C, Chen J. Decreasing survey response rates in the time of COVID-19: implications for analyses of population health and health inequities. Am J Public Health 2023;113(6):667-670. [doi: 10.2105/AJPH.2023.307267] [Medline: 37023386]
34. Vally Z. Public perceptions, anxiety and the perceived efficacy of health-protective behaviours to mitigate the spread of the SARS-Cov-2/ COVID-19 pandemic. Public Health 2020;187:67-73 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.002] [Medline: 32927291]
35. Malsam M. It's time to transition to inclusive personas. dscout People Nerds. 2024. URL: https://dscout.com/people-nerds/transition-personas [accessed 2024-07-15]
36. Polkosky M. When less is more: minimalist personas in UX. Medium. 2020. URL: https://uxdesign.cc/when-less-is-more-minimalist-personas-in-ux-f2ba742f0634 [accessed 2024-08-18]
37. Salminen J, Nielsen L, Jung S, An J, Kwak H, Jansen BJ. “Is More Better?”: Impact of multiple photos on perception of persona profiles. USA: Association for Computing Machinery; 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal QC Canada p. 1-13. [doi: 10.1145/3173574.3173891]
38. Pruitt J, Grudin J. Personas: practice and theory. USA: Association for Computing Machinery; 2003 Presented at: DUX '03: Proceedings of the 2003 conference on Designing for user experiences; June 7, 2003; San Francisco California p. 1-15. [doi: 10.1145/997078.997089]

Abbreviations

UCD: user-centered design
UW: University of Washington

Edited by E Mensah; submitted 21.Apr.2025; peer-reviewed by T Amusa, C Obianyo; comments to author 11.Jun.2025; revised version received 25.Oct.2025; accepted 18.Nov.2025; published 31.Mar.2026.

Please cite as:
Garcia JM, Hills RA, Revere D, Lim J, Elder AS, Baumgartner C, Karras BT, Baseman JG
Persona Development in Washington State: Mixed Methods Approach Using Statewide Survey Data
Online J Public Health Inform 2026;18:e75422
URL: https://ojphi.jmir.org/2026/1/e75422
doi:10.2196/75422
PMID:41915903

XSL•FO
RenderX

©Jordan M Garcia, Rebecca A Hills, Debra Revere, Jaewon Lim, Adam S Elder, Chris Baumgartner, Bryant T Karras, Janet G Baseman. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 31.Mar.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Instagram Memes of Oral Nicotine Pouches: Content Analysis Study

Samia Amin[1,2], PhD; Yunping Wu[2], BS; Samantha Baptista[2], BA; Scott K Okamoto[2], PhD; Thaddeus A Herzog[2], PhD; Crissy Terawaki Kawamoto[2], BSc; Pallav Pokhrel[2], PhD

[1]Department of Biology, Southern Arkansas University, Magnolia, AR, United States

[2]Population Sciences in the Pacific Program, University of Hawaii Cancer Center, 701 Ilalo Street, Honolulu, HI, United States

Corresponding Author:
Samia Amin, PhD
Department of Biology, Southern Arkansas University, Magnolia, AR, United States

Abstract

Background: Oral nicotine pouches (ONPs), such as Zyn, have gained popularity among young people; however, their portrayal on social media remains under-studied. Instagram memes, a widely shared form of digital communication, may shape young people's perceptions about ONPs and contribute to the widespread acceptance of ONP use.

Objective: This study examines the thematic content of Instagram memes related to ONPs to understand how these products are represented online.

Methods: The content of Instagram memes tagged with ONP-related hashtags—#oralnicotinepouch, #zyn, #on, #velo, and #nicotinepouch—was systematically analyzed. After screening, a total of 244 photo- and text-based memes were included in the final dataset. Using a structured coding framework, 3 researchers categorized the memes into key themes using NVivo software.

Results: Three dominant themes emerged: (1) the Zyn community (35.6%)—memes fostered a sense of belonging among users; (2) marketing and branding (27.8%)—humorous critiques of product advertising and accessibility; and (3) perceived consequences of use (13.9%)—memes highlighted perceived positive or negative consequences of ONP use. Engagement metrics revealed high levels of interaction, with the Zyn community theme garnering the most user engagement.

Conclusions: ONP-related Instagram memes are primarily focused on community identity, humor, and marketing, with community-centered content receiving the highest engagement. These findings indicate that social belonging and humor are central to the online representation of ONPs.

(Online J Public Health Inform 2026;18:e84025) doi:10.2196/84025

KEYWORDS

Instagram; memes; nicotine; oral nicotine pouches; tobacco; Zyn

Introduction

Oral nicotine pouches (ONPs) are a new type of tobacco product that, unlike traditional smokeless tobacco, use powdered nicotine rather than tobacco leaves [1]. They are typically used by placing them between the lip and gum, where they are held in the mouth to release nicotine [1]. The rising popularity of ONPs, particularly among young people, has drawn increasing attention from public health researchers and regulatory bodies [2]. Unlike traditional smokeless tobacco products, ONPs such as Zyn, On!, and Velo are marketed as tobacco-free alternatives that deliver nicotine without combustion, making them appealing to users seeking a discreet and convenient method of consumption [3]. However, short-term effects such as nausea, vomiting, bloating, and heartburn, along with long-term risks including nicotine addiction and potential oral health consequences, highlight the importance of understanding how these products are perceived in online environments [4,5]. These concerns are further amplified by the widespread perceptions of low harm of ONPs, influenced by their smokeless format and "tobacco-free" marketing [6]. Among various ONP brands, Zyn stands out as the most popular, widely recognized for its market dominance and consumer preference [7]. Despite the growing market presence of ONPs, at present, limited research has explored how ONPs are perceived and discussed within digital spaces, particularly through user-generated content on social media [7,8].

Most existing studies on ONPs have focused on marketing, branding, or sponsored content rather than user-generated discussions, leaving a gap in understanding how everyday users perceive and share these products online. On Reddit, prior research examined perceptions of ONPs through observational analysis of posts, revealing users' experiences and attitudes toward these products [9]. On TikTok, studies have analyzed ONP-related videos, highlighting their growing popularity, the promotional strategies used, and patterns of user engagement

XSL•FO
RenderX

[10-12]. Instagram (Meta Platforms, Inc), which has 3 billion monthly active users as of late 2025, serves as a key hub for the exchange of user-generated content, including nicotine consumption [13-16]. On Instagram, research has primarily focused on brand-marketed content, including visual advertisements and influencer posts illustrating how ONPs are promoted to large audiences [17]. Additional studies have examined ONP brand promotion through sponsorships and reward campaigns that encourage users to collect points for prizes; online retail tactics that leverage appealing product attributes such as flavors and brand reputation; and large-scale marketing campaigns across radio, television, and digital platforms, illustrating the diverse ways the ONP industry engages and interacts with consumers [18-20].

Research on Instagram posts more broadly shows how user-generated content can influence perceptions and social norms. Selfie-based portrayals on Instagram have demonstrated how smoking behaviors among young women can be glamorized through user-generated content, contributing to the normalization of tobacco use [15]. Similarly, Instagram posts related to e-cigarettes revealed that these products are framed as either harmful or socially acceptable [16]. Although prior research has examined ONP marketing content on Instagram [17], Instagram posts created by users, including memes, remain largely under-studied, representing an important gap in understanding peer-to-peer perceptions and social norms around ONP use.

Contents shared on Instagram as memes may especially influence perceptions and behavior related to nicotine products [21]. Memes are a widely shared and culturally relevant form of digital communication that usually uses wit and humor to further their content [22]. Young people often share tobacco- and marijuana-related memes with their friends, particularly memes that have received substantial attention in terms of likes and shares on social media [21]. However, how ONPs are depicted on Instagram memes remains unclear. As memes are an under-studied form of user-to-user communication, they can shape social norms, influence perceptions of product risk and acceptability, and provide early signals of emerging narratives that may inform prevention messages and policy-relevant communication strategies. Better understanding how the ONP-related content is represented in Instagram memes may provide valuable insights into the nature of the information about ONPs circulating on social media as well as potential social norms related to ONPs. To this end, this study seeks to conduct a systematic thematic content analysis of Instagram memes related to ONPs.

Methods

Ethical Considerations

This study was exempt from ethics review board approval because it involved the analysis of publicly available, anonymized data that did not include any identifiable private or sensitive personal information [23], consistent with previous tobacco-related social media research that has also been exempted from ethics approval for using public data [24,25].

Meme Identification

A systematic search was conducted on Instagram in September 2024 using the hashtags #oralnicotinepouch, #zyn, #on, #velo, and #nicotinepouch. We selected Instagram because it is one of the most commonly used social media platforms by young people, and memes are a prominent and widely shared content type on Instagram [22,26]. All data included were publicly accessible through Instagram's web interface. No private accounts or nonpublic posts were collected, ensuring compliance with Instagram's user agreement. Data were extracted manually through systematic searches on the Instagram platform; no application programming interface was used. Each post was documented and stored in a secure database with screenshots and URLs to ensure reproducibility.

Approximately 1800 posts across 8 Instagram accounts associated with #oralnicotinepouch, #zyn, #on, #velo, and #nicotinepouch were screened by the second and third authors. The inclusion criteria required the memes to be photo based and text based and originate from specific Instagram accounts associated with ONP-related hashtags. Posts were reviewed independently by the 2 coauthors to identify the eligible memes. Duplicate posts (n=47), videos (n=1086), posts from deleted accounts (n=396), or posts from accounts without meme content (n=27) were excluded. A total of 244 photo- and text-based memes met the inclusion criteria and were incorporated into the final database for analysis. We selected both photo- and text-based memes for analysis because doing so would enable us to examine both visual and textual elements, which was expected to provide a comprehensive understanding of how the message was communicated in terms of language, tone, and impact.

Codebook Development

The coding team consisted of 3 coauthors (YW, SB, and SA). Before coding commenced, the research team developed a detailed codebook. The team conducted a preparatory review of 10 randomly selected memes to familiarize themselves with the dataset and identify preliminary themes, visual styles, and recurring elements. The coding team collaborated using a structured approach to ensure consistency and accuracy in the analysis process.

For the development of the coding framework, the research team assigned each of the 3 coders (YW, SB, and SA) a subset of 30 randomly selected memes from the dataset. Each coder independently generated codes based on the primary themes and messages conveyed in the memes. Following this individual coding process, the team convened for a series of collaborative meetings to identify the most salient codes that aligned with the research objectives and grouped them into broader categories. Minor refinements were made to expand and remove a few codes based on team consensus, and all revisions were documented in an updated codebook to ensure transparency and auditability. The final codebook was organized into 3 major themes (Zyn community, marketing and branding, and perceived consequences of use), along with several additional but less prominent themes. A coding tree accompanied the codebook to ensure consistent application across the dataset. The codebook was developed primarily using an inductive approach; based

XSL•FO
RenderX

on a careful review of a sample of memes, no deductive coding based on existing frameworks was applied [27,28]. The unit of analysis was each individual Instagram post containing memes. For carousel posts, the entire carousel was treated as 1 post, and only the first meme image was coded, as it is the primary visual content displayed to users. The complete codebook with operational definitions for each code is provided as Multimedia Appendix 1.

The team used qualitative analysis software (NVivo; Lumivero) to code the memes systematically, with 2 team members (YW and SB) independently coding each meme. Team meetings were conducted regularly, both virtually and in person, to discuss progress and refine the codebook. Shared resources, including the dataset and coding materials, were organized and stored on Google Drive for seamless access and collaboration. Discrepancies were addressed through group discussions. Once consensus was reached, the codes were finalized and stored in a centralized repository for review and further analysis. The final codebook was used to systematically code the 244 memes.

Reliability and Validity

To ensure reliability, intercoder agreement using Cohen κ was periodically assessed throughout the coding process [29]. During pilot coding, Cohen κ ranged from 0.82 to 0.88 across all individual codes, and the mean κ of 0.85 (SD 0.02) was calculated for overall coding agreement, which demonstrated a high level of consistency and reproducibility in data categorization among all coders.

To maintain high construct validity, we ensured that the codes and categories accurately reflected the underlying concepts of health messaging, product promotion, and nicotine pouch use. This was done by obtaining continuous feedback from the research team and refining the codes based on the feedback [30].

Engagement Metrics Analysis

Engagement metrics were extracted for the 3 most prominent themes (Zyn community, marketing and branding, and perceived consequences of use) by collecting the total number of likes, comments, and shares associated with posts in each theme. Cumulative engagement values were calculated to quantify overall user interaction across themes.

Results

The majority of memes analyzed in this study were predominantly about Zyn, reflecting its strong presence on the Instagram platform. The analysis revealed a variety of themes with differing frequencies, reflecting the multifaceted nature of discussions surrounding ONP use.

Engagement Metrics for the Top 3 Prominent Themes

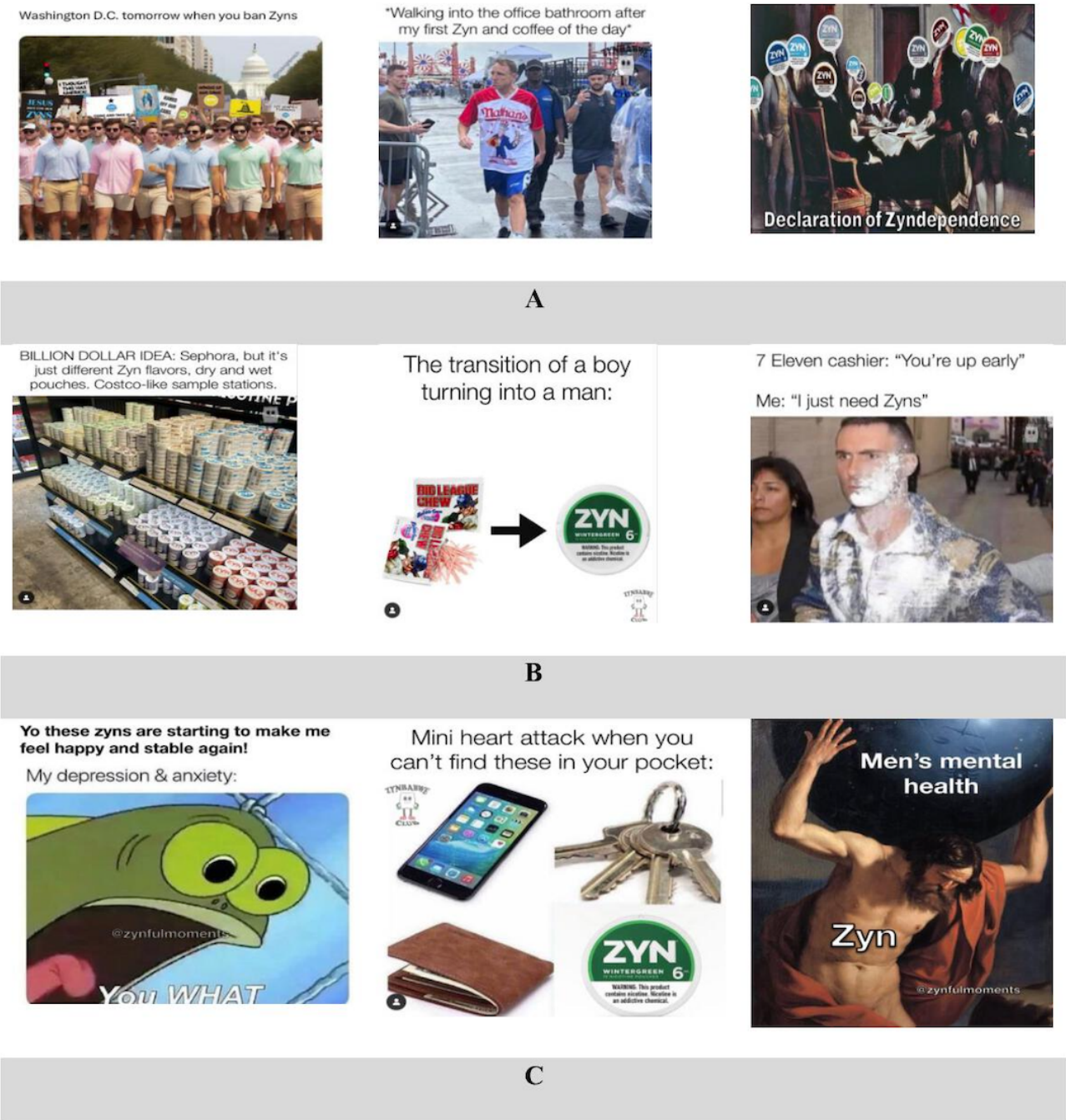
The engagement metrics revealed that the top 3 themes driving interaction were Zyn community, marketing and branding, and perceived consequences of use. The Zyn community category garnered significant interaction, with 650,000 likes, 1600 comments, and 696,000 shares. Similarly, the marketing and branding category achieved 585,000 likes, 738 comments, and 497,000 shares. In comparison, the perceived consequences of use category recorded 215,000 likes, 224,000 comments, and 422 shares.

Top 3 Prominent Themes

The most common meme category was Zyn community (87/244, 35.6%), followed by marketing and branding (68/244, 27.8%) and perceived consequences of use (34/244, 13.9%). Figure 1 shows a sample of memes that reflected these top 3 themes.

XSL•FO
RenderX

Figure 1. Memes that reflected these top 3 themes. (A) Sample memes illustrating the theme of Zyn community. (B) Sample memes illustrating the theme of marketing and branding. (C) Sample memes illustrating the theme of perceived consequences of use. Memes were posted by the following Instagram accounts, from left to right: (A), enhanced.snus [31], lippillowmaxxer [32], and the_zyn_head [33] (B), zyn_fluencers [34], zynbabweclub [35], zyncentral [36] (C), zynternet [37], enhanced.snus [31], zynfulmoments [38].



Theme 1: Zyn Community

The theme Zyn community highlighted the social aspect of Zyn use, where users developed group identity through their shared enthusiasm for the product. The memes in this category often combined humor with relatability to build a sense of belonging, allowing individuals to connect over their mutual use of Zyn. For example, a meme depicting people protesting in Washington, DC, after the Zyn ban tapped into the idea of collective defiance and solidarity within the Zyn user community. Another meme of a group of people staring at a man walking into the office bathroom after his first Zyn and coffee of the day highlights a sense of belonging among users who recognize and relate to the humorous scenario. Yet another meme showed a group of people facing Zyn and included the following text: "Declaration of Zyndependence." This meme seemed to highlight a sense of unity and shared pride among Zyn users. In sum, these memes appeared to foster bonding among Zyn users by highlighting shared experiences and in-group jokes related to Zyn.

Theme 2: Marketing and Branding

The marketing and branding theme primarily focused on how Zyn and other nicotine pouch brands positioned themselves in the market. Memes often humorously highlighted the flashy advertising strategies, product packaging, and brand slogans

XSL•FO
RenderX

used to appeal to consumers. For example, a meme stated, “Billion-dollar idea: Sephora, but it’s just a different Zyn flavor—dry and wet pouches. Costco sample station vibes,” thus humorously pointing out that Zyn’s marketing mimicked the aspirational branding strategies of high-end consumer goods. Another meme humorously showcased the transition from childhood to adulthood: “Big league chew to Zyn.” This meme reflects a generational shift from childhood to adulthood—referring to “Big League Chew,” a gum designed to mimic chewing Zyn, and portrays the transition to Zyn as a natural or expected part of growing up, potentially trivializing the shift to nicotine products. Other memes emphasized the ease of accessing Zyn and/or glorified the addiction to Zyn, for example, “7-Eleven cashier: You up early? Me: I just need Zyn.” Thus, these examples reflect how Zyn memes use humor to brand Zyn as an appealing product, glamorize Zyn use, and glorify the dependence on the product.

Theme 3: Perceived Consequences of Use

Several memes discussed the perceived consequences of using nicotine pouches, particularly in relation to mental health. One meme humorously stated, “Zyn are starting my day happy and stable again,” accompanied by an image of an animated character displaying a fearful, agitated expression, with the text “my depression and anxiety” displayed. Another meme humorously suggested that forgetting Zyn in one’s pocket—just like a phone or wallet—could cause a “mini heart attack,” emphasizing the intensity of dependence on the Zyn product. This exaggeration highlights how users may jokingly acknowledge their reliance on ONP products while continuing to engage with them.

Finally, a meme portrayed a man labeled as “Zyn” lifting a large boulder representing stress or negative affect. This meme seemed to communicate that the use of Zyn helped cope with high levels of stress or negative affect. Alternatively, the boulder may be interpreted to indicate high levels of dependence on Zyn. In sum, these memes reflect a mix of perceived positive and negative consequences associated with ONP use.

Additional Themes

Other themes identified in the analysis included nicotine strength, flavor varieties, first-time user experiences, and comparisons between Zyn and other tobacco product. When combined, these themes represent a minor portion of the overall dataset, contributing approximately 22.5% (55/244) of the total samples. Figure 2 shows a sample of memes that reflected these additional themes. Engagement metrics for other themes were not reported due to lower levels of interaction.

XSL•FO
RenderX

Figure 2. Memes that reflected additional themes. (A) Memes related to nicotine strength. (B) Memes related to flavor varieties. (C) Memes related to first-time user experiences. (D) Memes related to comparisons between Zyn and other tobacco products. Memes were posted by the following Instagram accounts: (A) lippillowmaxxer [32], (B) the_zyn_head [33], (C) zyncentral [36], and (D) zynternet [37].

XSL•FO
RenderX

Discussion

Principal Findings

This study explored ONP-related memes on the social media platform Instagram. The analysis revealed 3 prominent themes across the memes: Zyn community, marketing and branding, and perceived consequences of use. Memes in the Zyn community theme highlighted social bonding and shared experiences among users, while marketing and branding evaluated the potential appeal of ONPs. The perceived consequences of use theme captured the mixed perceptions of potential consequences of ONP use, such as stress or affect regulation and dependence. Additional themes, including nicotine strength, flavor varieties, first-time user experiences, and comparisons between Zyn and other tobacco products, were less prevalent.

Emerging tobacco products such as e-cigarettes are increasingly shaped by digital culture, where social media communities play a critical role in influencing how these products are perceived, shared, and normalized. Prior research on vaping and ONP-related social media content closely aligns with the Zyn community theme in this study, demonstrating how humor, shared narratives, and visually engaging posts foster a sense of belonging and normalize product use. Humor-driven content has been shown to strengthen community connections [39], while in-group communication and youth-friendly imagery help promote shared identity and normalize vaping behaviors [40]. Similarly, visually appealing posts such as selfies and product photos reinforce community bonds [41]. This pattern extends to ONPs, with Reddit discussions framing ONPs as lifestyle enhancers and TikTok content portraying them as trendy and youth-oriented through humorous, relatable narratives [9-11]. Our findings show that Instagram memes replicate these mechanisms, positioning ONPs as products that facilitate social connection and identity formation.

Advertisement, brand promotion, and marketing of e-cigarettes on social media have become prevalent practices [42-44]. Emerging evidence shows that ONPs are promoted on social media in ways similar to other tobacco products, such as e-cigarettes, with content framing them as socially acceptable, lifestyle-oriented, and often humorous, appealing to younger audiences [9-11]. Studies have shown that images depicting vaping activities tend to generate high engagement and that platforms such as Instagram and Pinterest are effective channels for e-cigarette product displays [11,42]. These findings align with our study's marketing and branding theme, which highlights Zyn's marketing strategies, particularly through memes that humorously position the brand as aspirational and ubiquitous, while also suggesting that Zyn use is a grown-up behavior. Research studies also indicate that ONPs are marketed for noncessation purposes and commonly use youth-targeted cues, lifestyle imagery, and implicit health messaging [12,17]. Collectively, these findings reinforce that ONP marketing on social media, much like e-cigarette marketing, leverages visual storytelling, humor, and lifestyle branding to broaden appeal, particularly among younger and digitally engaged audiences.

The perceived consequences of use theme reflects the mixed beliefs surrounding ONP use. Consistent with recent TikTok and Reddit studies, our findings show that users often frame ONPs as tools for stress relief or affect regulation, despite contradictory evidence [9,10]. Memes in this category frequently portray Zyn as a coping mechanism for stress, affect, or poor mental health symptoms (eg, anxiety and depression) [11] or a humorous dependency (eg, the "mini heart attack" when a pouch is missing). These portrayals align with findings from nicotine research indicating that users often perceive nicotine as stress reducing, although nicotine can exacerbate stress and withdrawal symptoms [45]. The humorous framing in memes may obscure these harms, reinforcing the normalization and trivialization of dependence.

The findings from this study highlight significant implications for public health strategies targeting ONP use. First, given the widespread use of humor and relatability in memes, public health campaigns could consider adopting similar strategies to effectively communicate the health risks associated with ONPs. By leveraging humor, these campaigns may better engage younger and more digitally connected audiences, potentially reshaping the cultural narratives surrounding nicotine use. Second, the study emphasizes the need for policies that address the normalization of these products in online spaces. Regulatory agencies should recognize memes and other user-generated content as influential promotional tools and consider monitoring these platforms, as understanding how memes reinforce marketing tactics and community identities can inform regulatory frameworks aimed at controlling the promotion and accessibility of these products, particularly on youth-oriented digital platforms. At the same time, public health professionals have an opportunity to engage with active communities, such as the online Zyn community on Instagram, using their language and humor to provide accurate, balanced, and reliable information about nicotine, including both addiction risks and potential harm-reduction aspects of noncombustible products. Third, policies should also target indirect marketing practices, including influencer partnerships, trend-based content, and meme-driven brand symbolism, which often escape traditional advertising oversight. By combining engaging public health messaging with updated regulatory measures, it may be possible to reduce the appeal of ONPs among adolescents and young adults.

Strengths, Limitations, and Future Directions

This study has several strengths that enhance its relevance to tobacco research. First, it investigates the growing online community of ONP users, using memes that capture the social dynamics and cultural narratives surrounding these products. The analysis also examines how memes blend humor and social commentary, offering insights into user perceptions and behaviors. This is particularly relevant to young people—a group especially vulnerable to nicotine initiation and addiction—as such content may reinforce interest in and normalize nicotine pouch use. By focusing on the popular social media platform Instagram, this study highlights the need for innovative public health interventions that resonate with digital communities. Furthermore, the research sheds light on the influence of user-generated content in shaping product

XSL•FO
RenderX

perceptions and cultural identities, contributing to the broader conversation about how digital platforms impact health behaviors.

While our study offers valuable insights into the social and cultural dimensions of ONP use, there are several limitations to consider. The analysis was limited to Instagram, which was selected due to its popularity among young adults and its prominence in visual meme-based content. However, nicotine-related messaging also appears on platforms such as TikTok, X (formerly Twitter), and Reddit, which were not examined here. The reliance on memes as a primary data source may also have introduced skewed perceptions, as memes often exaggerate or trivialize experiences for comedic effect and may not fully capture the complexities of user experiences or broader attitudes. The hashtag-based sampling strategy may have excluded relevant content outside those parameters and increased the likelihood of including posts from nonhuman accounts, such as automated promotional profiles. Additionally, posts were manually screened, and lower-engagement content was removed, which limits reproducibility and generalizability and may have omitted important perspectives. Finally, the engagement matrix was reported as cumulative totals per theme, which may limit variation at the individual post level.

Future research should explore cross-platform differences in nicotine-related discourse to develop a more comprehensive understanding of how ONPs are represented and circulated online. Combining meme analysis with traditional qualitative methods, such as focus groups or in-depth interviews with members of digital communities, could provide a more nuanced understanding of how ONP use is perceived across different demographic groups. Incorporating bot detection tools and broader sampling strategies could help improve the rigor of future analyses. Studies using post-level engagement data and applying descriptive or inferential statistical techniques may more accurately characterize patterns of user interaction. Additionally, longitudinal designs could help track evolving cultural perceptions of nicotine products as they become more mainstream, and future work may also examine how memes contribute to the adoption of other emerging nicotine products, such as smokeless and heated tobacco products.

Conclusions

Our study illustrates how memes reflect the perceptions of ONPs, emphasizing key themes of community, marketing, and consequences of use. The findings highlight the importance of understanding the digital culture surrounding these products, particularly as they gain popularity among younger, more digitally connected populations. By recognizing the role of humor, relatability, and branding in influencing attitudes, public health campaigns can be better tailored to address the cultural appeal of ONPs and mitigate their potential harm and understand their potential as substitutes for cigarettes. The study lays the groundwork for future research that can explore the intersection of digital culture, substance use, and public health, offering valuable insights for designing more effective, culturally relevant health interventions and providing reliable, balanced, and evidence-based information on nicotine and tobacco products.

Funding

This research was supported by grant from National Cancer Institute: R25CA244073.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comprehensive hierarchical codebook detailing themes, operational definitions, indicators, and example applications used to systematically guide and standardize the qualitative analysis.
[PDF File, 140 KB - ojphi_v18i1e84025_app1.pdf]

References

1. Robichaud MO, Seidenberg AB, Byron MJ. Tobacco companies introduce “tobacco-free” nicotine pouches. Tob Control 2020 Dec;29(e1):e145-e146. [doi: 10.1136/tobaccocontrol-2019-055321] [Medline: 31753961]
2. Patel M, Kierstead EC, Kreslake J, Schillo BA. Patterns of oral nicotine pouch use among U.S. adolescents and young adults. Prev Med Rep 2023 Aug;34:102239. [doi: 10.1016/j.pmedr.2023.102239] [Medline: 37228836]
3. Patwardhan S, Fagerström K. The new nicotine pouch category: a tobacco harm reduction tool? Nicotine Tob Res 2022 Mar 1;24(4):623-625. [doi: 10.1093/ntr/ntab198] [Medline: 34606613]
4. Al-Nafisi AM, Alsulami ZB, Alsulami MB, et al. Assessing the gastrointestinal and psychological impacts of nicotine pouch use among adults in Saudi Arabia: a cross-sectional study. Tob Prev Cessat 2025;11:10. [doi: 10.18332/tpc/207753] [Medline: 40831452]
5. Reimann H, Berger M, Eckert E, Merches K, Börnke F. Beyond smoking: risk assessment of nicotine in pouches. Toxicol Rep 2024 Dec;13:101779. [doi: 10.1016/j.toxrep.2024.101779] [Medline: 39554607]

XSL•FO
RenderX

6. Miluna-Meldere S, Vanka SA, Skadins I, Kroica J, Sperga M, Rostoka D. Oral mucosal changes caused by nicotine pouches: case series. Diagn Pathol 2024 Sep 19;19(1):127. [doi: 10.1186/s13000-024-01549-3] [Medline: 39300504]
7. Liu J, Halpern-Felsher B. Popularity of Zyn and other nicotine pouch brands among U.S. adolescents, young adults, and adults in 2021. Prev Med 2024 Aug;185:108068. [doi: 10.1016/j.ypmed.2024.108068] [Medline: 39002807]
8. Majmundar A, Okitondo C, Xue A, Asare S, Bandi P, Nargis N. Nicotine pouch sales trends in the US by volume and nicotine concentration levels from 2019 to 2022. JAMA Netw Open 2022 Nov 1;5(11):e2242235. [doi: 10.1001/jamanetworkopen.2022.42235] [Medline: 36378312]
9. Shao Y, Zou J, Xie Z, et al. Perceptions of oral nicotine pouches on Reddit: observational study. J Med Internet Res 2022 Jul 15;24(7):e37071. [doi: 10.2196/37071] [Medline: 35838764]
10. Mand A, Fonteyne K, Struik L. Examining how oral nicotine pouches are trending on TikTok: a qualitative descriptive study. JMIR Form Res 2025 Nov 14;9:e73032. [doi: 10.2196/73032] [Medline: 41236804]
11. Donaldson SI, La Capria K, DeJesus A, Ganz O, Delnevo CD, Allem JP. Describing ZYN-related content on TikTok: content analysis. Nicotine Tob Res 2026 Jan 22;28(2):277-281. [doi: 10.1093/ntr/ntaf016] [Medline: 39823221]
12. Zenone M, Harries B, Hartwell G. The promotion of oral nicotine pouches for non-smoking cessation purposes on TikTok. Nicotine Tob Res 2026 Jan 22;28(2):282-286. [doi: 10.1093/ntr/ntaf024] [Medline: 39879071]
13. Number of Instagram users worldwide from 2020 to 2025. Statista. 2025. URL: https://www.statista.com/statistics/183585/instagram-number-of-global-users/ [accessed 2025-12-13]
14. Casaló LV, Flavián C, Ibáñez-Sánchez S. Be creative, my friend! Engaging users on Instagram by promoting positive emotions. J Bus Res 2021 Jun;130:416-425. [doi: 10.1016/j.jbusres.2020.02.014]
15. Cortese DK, Szczypka G, Emery S, Wang S, Hair E, Vallone D. Smoking selfies: using Instagram to explore young women's smoking behaviors. Soc Media Soc 2018 Jul;4(3). [doi: 10.1177/2056305118790762]
16. Lim YS, Lee JY. A comparative analysis of E-cigarette and cigarette posts on Instagram. Int J Environ Res Public Health 2023 Feb 10;20(4):3116. [doi: 10.3390/ijerph20043116] [Medline: 36833814]
17. Czaplicki L, Tfayli D, Spindle TR, et al. Content analysis of marketing features in US nicotine pouch ads from 2021 to 2023. Tob Control 2024 Dec 3:tc-2024-059010. [doi: 10.1136/tc-2024-059010] [Medline: 39627024]
18. Dobbs PD, Kong G, Berman ML, Henriksen L. "Cashing in" nicotine pouches for prizes. Tob Control 2024 Jun 15:tc-2024-058691. [doi: 10.1136/tc-2024-058691] [Medline: 38879181]
19. Hardin A, Ahn S, Ma S, et al. The association between product attributes and oral nicotine pouch preference—evidence from online stores. Nicotine Tob Res 2026 Jan 22;28(2):298-303. [doi: 10.1093/ntr/ntaf137] [Medline: 40590502]
20. Duan Z, Henriksen L, Vallone D, et al. Nicotine pouch marketing strategies in the USA: an analysis of Zyn, On! and Velo. Tob Control 2024 Mar;33(2):154-163. [doi: 10.1136/tc-2022-057360]
21. Liu J, McLaughlin S, Lazaro A, Halpern-Felsher B. What does it meme? A qualitative analysis of adolescents' perceptions of tobacco and marijuana messaging. Public Health Rep 2020;135(5):578-586. [doi: 10.1177/0033354920947399] [Medline: 32791026]
22. Dixon S. Instagram: distribution of global audiences 2024, by age group. Statista. 2024. URL: https://www.statista.com/statistics/325587/instagram-global-age-group/ [accessed 2025-12-13]
23. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):708-713. [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
24. Hardie L, Watts C, Freeman B, McCool J. An analysis of social media strategies used to promote the E-Cigarette brand Vuse. Health Promot Int 2025 Jul 1;40(4):daaf114. [doi: 10.1093/heapro/daaf114] [Medline: 40795924]
25. Amin S, Jaiswal A, Washington PY, Pokhrel P. Investigating #vapingcessation in Twitter. am J health behav 2023 Dec 31;47(6):1183-1191. [doi: 10.5993/AJHB.47.6.11]
26. Lorenz T. Text memes are taking over Instagram. The New York Times. 2021. URL: https://www.nytimes.com/2021/08/09/technology/instagram-text-memes.html [accessed 2025-12-13]
27. Bingham AJ. From data management to actionable findings: a five-phase process of qualitative data analysis. Int J Qual Methods 2023 Oct;22. [doi: 10.1177/16094069231183620]
28. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772. [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
29. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 1;19. [doi: 10.1177/1609406919899220]
30. Clark LA, Watson D. Constructing validity: new developments in creating objective measuring instruments. Psychol Assess 2019 Dec;31(12):1412-1427. [doi: 10.1037/pas0000626] [Medline: 30896212]
31. Enhanced.snus Instagram profile. Instagram. URL: www.instagram.com/enhanced.snus [accessed 2024-9-03]
32. Lippillowmaxxer Instagram profile. Instagram. URL: www.instagram.com/lippillowmaxxer [accessed 2024-09-05]
33. The_zyn_head Instagram profile. Instagram. URL: www.instagram.com/the_zyn_head [accessed 2024-09-07]
34. Zyn_fluencers Instagram profile. Instagram. URL: www.instagram.com/zyn_fluencers [accessed 2024-09-10]
35. Zynbabweclub Instagram profile. Instagram. URL: www.instagram.com/zynbabweclub [accessed 2024-09-14]
36. Zyncentral Instagram profile. Instagram. URL: www.instagram.com/zyncentral [accessed 2024-09-18]

XSL•FO
RenderX

37. Zynternet Instagram profile. Instagram. URL: www.instagram.com/zynternet [accessed 2024-09-22]
38. Zynfulmoments Instagram profile. Instagram. URL: www.instagram.com/zynfulmoments [accessed 2024-09-26]
39. Basch CH, Fera J, Pellicane A, Basch CE. Videos with the hashtag #vaping on TikTok and implications for informed decision-making by adolescents: descriptive study. JMIR Pediatr Parent 2021 Oct 25;4(4):e30681. [doi: 10.2196/30681] [Medline: 34694231]
40. Majmundar A, Pérez C, Huerta M, Unger JB, Allem JP. Describing memes referencing vaping: thematic analysis. Subst Use Misuse 2023;58(2):306-310. [doi: 10.1080/10826084.2022.2161316] [Medline: 36585016]
41. Ketonen V, Malik A. Characterizing vaping posts on Instagram by using unsupervised machine learning. Int J Med Inform 2020 Sep;141:104223. [doi: 10.1016/j.ijmedinf.2020.104223] [Medline: 32623330]
42. Chu KH, Allem JP, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Sep;26(5):575-578. [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
43. Chu KH, Matheny SJ, Sidani JE, Allem JP, Unger JB, Cruz TB. Instagram's #JUUL: who's posting what. Transl Behav Med 2021 Feb 11;11(1):257-261. [doi: 10.1093/tbm/ibz169] [Medline: 31912873]
44. Lee AS, Hart JL, Sears CG, Walker KL, Siu A, Smith C. A picture is worth a thousand words: electronic cigarette content on Instagram and Pinterest. Tob Prev Cessat 2017 Jul;3:119. [doi: 10.18332/tpc/74709] [Medline: 28815224]
45. Bares CB, Dick DM, Kendler KS. Nicotine dependence, internalizing symptoms, mood variability and daily tobacco use among young adult smokers. Addict Behav 2018 Aug;83:87-94. [doi: 10.1016/j.addbeh.2017.09.004] [Medline: 28943065]

Abbreviations

ONP: oral nicotine pouch

Edited by EK Achampong; submitted 12.Sep.2025; peer-reviewed by JF Etter, KL Capria, P Dobbs; revised version received 13.Dec.2025; accepted 19.Dec.2025; published 24.Mar.2026.

Please cite as:
Amin S, Wu Y, Baptista S, Okamoto SK, Herzog TA, Kawamoto CT, Pokhrel P
Instagram Memes of Oral Nicotine Pouches: Content Analysis Study
Online J Public Health Inform 2026;18:e84025
URL: https://ojphi.jmir.org/2026/1/e84025
doi:10.2196/84025

© Samia Amin, Yunping Wu, Samantha Baptista, Scott K Okamoto, Thaddeus A Herzog, Crissy Terawaki Kawamoto, Pallav Pokhrel. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 24.Mar.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessment of the Cultural Nuances in COVID-19 Vaccine Uptake Through a Comparative Analysis of English and Spanish Facebook Posts in Tarrant County, Texas: Longitudinal Study

Ana Aleksandric[1,2], PhD; Anisha Dangal[3], MPH; Shirin Nilizadeh[2], PhD; Gabriela Mustata Wilson[3,4], PhD

[1]Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, United States

[2]Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington, TX, United States

[3]Department of Kinesiology, The University of Texas at Arlington, 701 S Nedderman Dr, Arlington, TX, United States

[4]Louisiana Center for Health Innovation, University of Louisiana at Lafayette, 635 Cajundome Blvd., Lafayette, LA, United States

Corresponding Author:
Gabriela Mustata Wilson, PhD
Department of Kinesiology, The University of Texas at Arlington, 701 S Nedderman Dr, Arlington, TX, United States

Abstract

Background: Prior studies have identified key factors contributing to COVID-19 vaccine hesitancy, including concerns over vaccine safety, potential side effects, and mistrust in the health care system. According to the World Health Organization, vaccine hesitancy is among the top 10 threats to global public health. Previous research has suggested that vaccine hesitancy is a significant barrier within the Hispanic population, particularly in Texas.

Objective: This longitudinal study examined the relationships of daily stance, misinformation, and topics in vaccine-related English and Spanish Facebook posts with daily vaccination rates in Tarrant County, Texas, during 2021 and 2022. The goal was to identify the predictors associated with vaccination uptake and inform targeted social media interventions, with particular attention to the Hispanic population.

Methods: COVID-19 vaccine-related English and Spanish Facebook posts from Tarrant County were collected for 2021 and 2022. The study analyzed 12,395 English posts and 1123 Spanish posts. Posts were annotated using GPT-4 for stance, misinformation, and relevant topics, including vaccine availability, safety, and side effects. Category prevalence was compared across English and Spanish posts and across years. Linear regression models were used to examine associations between post characteristics and daily vaccination rates in the total and Hispanic populations.

Results: Regression analysis identified distinct predictors of Hispanic vaccination uptake, including encouraging posts (P=.02) and religion-related posts (P=.007), which were not significant predictors for vaccination uptake in the general population. A substantial proportion of Spanish discouraging posts focused on vaccine side effects (13/70, 19%) and health system distrust (24/70, 34%), suggesting concerns that may be especially relevant within the Hispanic community. Predictors associated with higher uptake in both the Hispanic and total populations included posts related to vaccine availability (P=.01), vaccine safety (P=.006), and misinformation debunking (P<.001).

Conclusions: Posts related to vaccine availability, vaccine safety, and debunking misinformation were associated with higher vaccination uptake. Encouraging posts and religion-related posts were associated with higher vaccination uptake in the Hispanic population, suggesting meaningful cultural nuances. These findings support the value of culturally tailored social media messaging in public health campaigns.

(Online J Public Health Inform 2026;18:e72465) doi:10.2196/72465

KEYWORDS

social media; COVID-19 vaccine hesitancy; cultural nuances; Hispanic population; misinformation; public health interventions; vaccination acceptance

Introduction

Background

In March 2020, the World Health Organization (WHO) declared COVID-19 a pandemic, underscoring the critical need for effective public health measures, including vaccination, to control the spread of the virus [1]. Vaccination has historically been one of the most effective interventions in preventing infectious diseases and protecting population health [2]. Despite the availability and widespread administration of COVID-19 vaccines, many individuals remain hesitant to receive them, raising serious public health concerns [3]. WHO identified vaccine hesitancy as one of the top 10 global health threats in

XSL•FO
RenderX

2019, even before the onset of COVID-19 [4]. This persistent hesitancy has been driven by factors, including vaccine safety concerns, fears of side effects [1], mistrust in the health care system, various socioeconomic and demographic influences [5,6], and the usage of social media platforms for gathering information [7,8].

These factors do not affect all communities equally. Prior research has shown higher vaccine hesitancy in some ethnic minority populations [9]. Studies examining racial and ethnic disparities in COVID-19 vaccination found that people with lower income or education, women, and some minority groups, including Black and Hispanic populations, were more likely to delay or avoid vaccination [10,11]. Such findings underscore the importance of localized analysis, since state- and national-level patterns may mask community-specific concerns.

This is particularly relevant in Texas. During the 2021 legislative session, vaccine hesitancy discussions reflected concerns about medical freedom, vaccine effectiveness, and vaccine safety [12]. While these concerns overlap with national patterns, they may not fully capture barriers affecting smaller communities, including Hispanic populations. National-level data have shown that Hispanic individuals were less likely than White individuals to receive an initial vaccine dose [13]. Other research has suggested that messaging emphasizing family protection and directly addressing mistrust may be more effective for Latino communities [14]. Together, these findings suggest that interventions may need to be tailored to not only demographic groups but also local settings.

Social media has been widely used to study COVID-19 vaccine hesitancy [15-18]. Some studies focused on sentiment [19-25], whereas others examined misinformation [16,26,27]. Prior work has analyzed English-only [28,29], Spanish-only [30-33], and multilingual datasets [34-37]. For example, one study examined vaccine misinformation in English and Spanish [34], and another showed that multilingual discussion around AstraZeneca and Omicron underscored the need for culturally informed public messaging [35]. However, prior research has rarely linked bilingual social media discourse to real-world vaccination uptake at the county level.

This study addresses that gap by analyzing English and Spanish Facebook posts from public pages and groups in Tarrant County, Texas, across 2021 and 2022. Tarrant County is a large, diverse urban county with a substantial Hispanic population (30.5%) [38], making it a useful setting for examining cultural influences on vaccine uptake. By combining bilingual social media data with vaccination data disaggregated by time and ethnicity, this study provides a more granular view of how online discourse relates to population-level vaccination behavior.

The study addresses three research questions:

1. How did stance, misinformation, and topics differ between English and Spanish social media posts regarding COVID-19 vaccines in Tarrant County?
2. How did these differences in stance, misinformation, and topics change over time from 2021 to 2022?
3. What associations existed between the characteristics of these posts and daily vaccination rates among the total and Hispanic populations?

This analysis is informed by two complementary frameworks: (1) the health belief model [39] and (2) the agenda-setting theory [40]. The health belief model suggests that health behaviors are shaped by perceived risk, benefits, barriers, and cues to action [39]. In this context, social media posts may act as cues to action by shaping how people think about vaccine safety, effectiveness, and accessibility. The agenda-setting theory emphasizes the role of media in influencing which issues are seen as important [40]. Applied here, it suggests that frequent discussion of issues, such as side effects, mandates, and misinformation, may increase their salience and thereby shape public attitudes and behaviors. Together, these frameworks provide a useful basis for understanding how vaccine-related discourse on Facebook may be associated with vaccine uptake.

In collaboration with the Tarrant County Public Health Department, this study used monthly vaccination records that included ethnicity, enabling the analysis of both total and Hispanic vaccination uptake. The study also introduced a scalable approach for annotating vaccine-related posts using GPT-4. Post-level variables were defined through iterative manual coding of 200 English and 200 Spanish posts, and these were then used to construct prompts for labeling the full dataset across 16 variables. By examining stance, misinformation, and a wide range of vaccine-related topics, this study has identified message types associated with higher vaccination uptake and highlighted cultural differences that may help guide future public health communication.

Variables

Overview

The analysis included dependent, independent, and control variables. Because Facebook posts may contain more than one relevant characteristic, independent variables were not mutually exclusive.

Dependent Variables

Vaccination data from January 2021 to December 2022 were obtained from the Tarrant County Public Health Department. The data were anonymized. Each data point represented an individual who got vaccinated and contained information such as the total number of doses received, race, ethnicity, gender, date of birth, and the date of the last dose. It is important to note that these data are not publicly available, and merging these data with social media posts made this research possible. The data were received monthly, where new individuals and doses were added to the record. Such data allowed for the calculation of the total new daily vaccinations at the county level and the new daily vaccinations of the Hispanic population at the county level. New daily vaccinations represent new individuals who have received their first dose that day. Thus, the analysis used new vaccinations per day as a dependent variable. In summary, we defined the dependent variables *new Hispanic vaccinations* and *new total vaccinations* as newly vaccinated Hispanic individuals per day and newly vaccinated individuals in total per day, respectively. Separating the total vaccination rate from

XSL•FO
RenderX

the vaccination rate of the Hispanic population was a crucial step in understanding cultural nuances in vaccination uptake, as it allowed the examination of which factors are significant predictors compared with total vaccination uptake. It is important to note that the dependent variables (ie, *new Hispanic vaccinations* and *new total vaccinations*) were calculated by dividing the daily number of first-dose COVID-19 vaccinations by the corresponding population groups in Tarrant County, Texas. Specifically, *new Hispanic vaccinations* were normalized using the county's Hispanic population, while *new total vaccinations* were normalized using the county's total population. This standardization (ie, vaccinations per capita) accounts for differences in population size and enables more accurate comparisons over time and between demographic groups.

Independent Variables

In the literature, stance is defined as an individual's standpoint toward a proposition or topic (eg, abortion, feminism, climate change, and others), where the viewpoint can be supporting, opposing, or neutral toward the object [41]. Therefore, the concept of stance could be applied as the users' standpoint toward vaccination, formulating the first independent variable.

According to multiple research studies, becoming acquainted with antivaccination views and false information on Twitter contributed to vaccine reluctance and denial as well as an overall decrease in the number of individuals receiving the vaccine [42-44]. Thus, *vaccination stance* was used as an independent variable to examine its possible association with the vaccination rate. Encouraging and discouraging posts were included, while neutral posts were not considered. As a neutral stance does not provide any specific insights and does not reflect users' views toward vaccination, its exclusion allows for a clearer comparison of the associations of encouraging and discouraging posts with the vaccination rate. In more detail, encouraging posts are posts that explicitly support vaccination and aim to motivate others to get vaccinated, while discouraging posts express a negative standpoint toward vaccination and aim to prevent vaccination.

Prior work defined misinformation as unintentionally shared information that is fake or misleading [45]. According to related research, misinformation propagated online hurts COVID-19 vaccine acceptance [46-50]. Some of the most common conspiracy theories regarding COVID-19 vaccines were that they do not work and that they may cause autism or infertility [46]. In addition, with many social media platforms nowadays, there is greater exposure to vaccine misinformation [51], potentially slowing progress toward reaching herd immunity [26]. However, individuals with higher vaccine literacy are less likely to believe in such theories [52,53]. Hence, statistical models used vaccine *misinformation* as an independent variable. Similar to the approach used for stance, only debunking and misinformation posts were included, while posts without relevant content were not considered, as they do not provide useful insights.

Previous research measuring the effectiveness of interventions aiming to reduce vaccine hesitancy toward influenza, human papillomavirus, tetanus, polio, and other vaccines concluded that most interventions included multiple aspects that focus on raising knowledge and vaccine awareness [54]. Many recent studies have proposed the use of social media platforms for targeted messaging to educate the population and enhance awareness in order to reduce vaccine hesitancy [29,46]. Therefore, it is evident that spreading accurate information regarding vaccination can increase vaccination acceptance, and social media could aid in outreach. Thus, it is important to consider which social media posts provide factual information regarding vaccination. Whether a post is *informative* was used as a binary independent variable in the models.

There are several *post topics/categories*. The literature has highlighted multiple reasons behind vaccine hesitancy, such as potential vaccine side effects, mistrust toward the government or health system, doubts about vaccine benefits/efficacy and safety, and dissatisfaction with vaccination mandates [1,5,6,55]. Thus, topics, such as *vaccine side effects*, *government*, *health system*, *vaccine safety*, *policies/mandates*, and *vaccine benefits/efficacy*, were included as independent binary variables in the statistical models. Furthermore, prior work has shown that study participants who knew someone who contracted COVID-19 or died because of it were more likely to receive the vaccine [56]. Therefore, the variable indicating whether a post discusses a COVID-19 illness experience was included as an independent variable. Research has also suggested that religious beliefs play a significant role in vaccine hesitancy, where religiosity is negatively associated with COVID-19 vaccine uptake [57]. Thus, whether a post mentions any religious beliefs was used as a binary independent variable in the models. Finally, vaccine hesitancy is more prevalent across minority populations [14,58]. Therefore, *community-specific advice* variables indicating whether a post includes information specific to any minority population were included as binary variables in the analysis.

While certain variables were identified as independent based on previous literature, others were defined through open-coding techniques during the data labeling process. During the annotation of Facebook posts, annotators identified several frequently occurring categories that were subsequently added as independent variables in the models owing to their potential influence on vaccine uptake. These categories—*vaccine availability*, *education*, *postvaccination advice*, and *statistics*—represent critical aspects of public discourse that address practical, emotional, and informational barriers to vaccination. Vaccine availability posts provide essential details on where, when, and how to get vaccinated, eliminating logistical uncertainties hindering vaccine uptake. Posts related to education discuss school policies and vaccination requirements, which are particularly relevant for parents and students navigating the return to in-person learning. Postvaccination advice posts offer guidance on managing side effects and follow-up doses, addressing anxieties, and simplifying the vaccination process. Finally, posts sharing statistics present concrete data on vaccination rates and COVID-19 cases, serving as powerful motivators by reinforcing the urgency and benefits of vaccination. Including these variables in the models allowed for a more comprehensive analysis of how different types of messaging influence

XSL•FO
RenderX

vaccination behavior, providing actionable insights for designing effective, targeted public health interventions.

Control Variables

The analysis included a measure of the *population available to vaccinate*, calculated as the share of the relevant population that has not yet received a vaccine. Separate measures were used for the total and Hispanic populations. This variable accounts for the declining pool of eligible first-dose recipients over time.

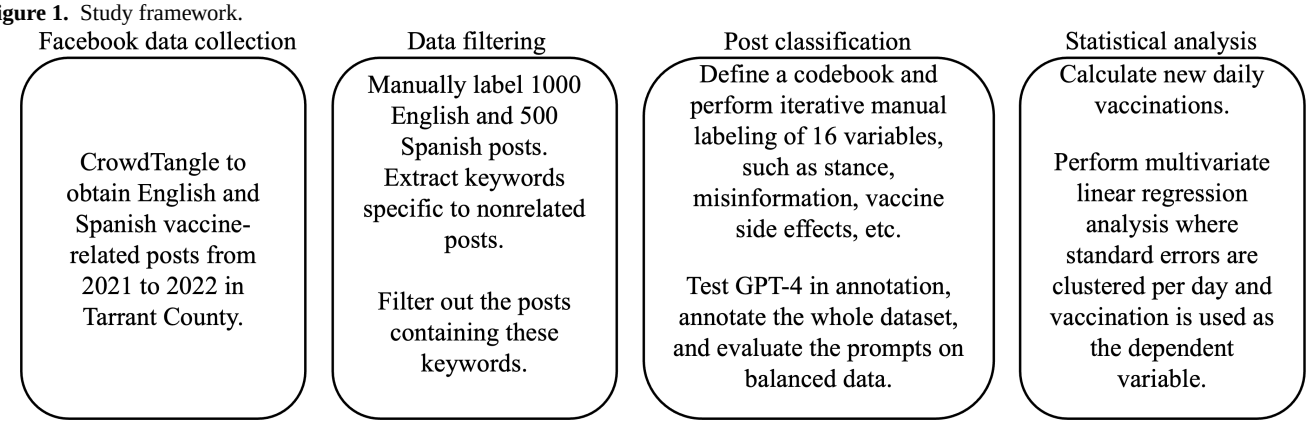
Language was also included as a control variable coded as English or Spanish. This allowed the analysis to test whether post language was associated with differential vaccination uptake while recognizing that language does not map perfectly onto ethnicity.

Methods

Overview

Figure 1 presents the study framework, which included the following 4 stages: Facebook data collection, data filtering, post classification, and statistical analysis.

Figure 1. Study framework.



Facebook Data Collection

Many studies have leveraged social media platforms to assess vaccine hesitancy and vaccination acceptance. For example, studies have examined sentiments and misinformation about vaccines on platforms like Twitter, Facebook, YouTube, and WhatsApp [29,59-61]. This study collected data from one of the most widely used social media platforms, Facebook [62]. Facebook was considered a suitable platform for this study, as a previous study has reported that 66% of the Hispanic population uses this platform [63]. CrowdTangle application programming interface [64], a social media listening tool from Meta, was leveraged to query the data based on the selected keywords, location (Tarrant County), and period (January 1, 2021, to December 31, 2022). It is important to note that the data provided by CrowdTangle originate from public pages and public groups, not individual users, while the selected location yields posts relevant to that region based on the location distribution of page/group followers or participants. The keywords used for collecting the English dataset were adopted from the CoVaxxy dataset [65], which is an extensive collection of COVID-19 vaccine-related tweets collected with a unique keyword list created by using the snowballing technique. Some examples include *pfizervaccine*, *covax*, and *getvaccinated*. The keywords for collecting the Spanish dataset were adopted from another related study analyzing Spanish Facebook posts in Texas [30], where the keyword list was also created by incorporating a snowballing technique to identify relevant keywords specific to Spanish COVID-19 vaccine-related posts. Some examples include *vacuna covid*, *segunda dosis*, and *efecto secundario covid*. Finally, 21,737 English posts and 1596 Spanish posts were collected. There may be a few reasons for the lower number of Spanish posts. First, the percentage of the Hispanic population in Tarrant County is around 30%. Therefore, the percentage of Spanish versus English posts is expected to be lower. Second, some Hispanic users might still post in English. While this study aimed to compare social media post categories between Hispanic and non-Hispanic populations, a key limitation is the inability to definitively determine the ethnicity of users based on post language. The assumption that Spanish-language posts originate from Hispanic individuals and English-language posts originate from non-Hispanic users may not always be accurate. This introduces potential biases in the analysis, as users may post in a language that does not reflect their ethnicity. This limitation affects the validity of the conclusions drawn about how different demographic groups express attitudes toward vaccination. Despite this, the findings remain relevant because they identify key post categories associated with vaccination uptake in each language group. These insights suggest that the sample is still broadly representative, even if individual user ethnicity cannot be verified with certainty.

Ethical Considerations

The study was reviewed by the University of Texas at Arlington Institutional Review Board (protocol #2022 - 0072), which determined that it did not meet the definition of human subjects research under 45 CFR 46.104 (d)(2i), Revised 2018 [66]. As such, the study was classified as non–human subjects research and did not require Institutional Review Board approval. The project used publicly available data and was consistent with public health surveillance activities, as it involved the systematic collection and analysis of population-level information to monitor health-related behaviors and information patterns, without interaction with individuals or access to identifiable private information, which has been outlined in Centers for

XSL•FO
RenderX

Disease Control and Prevention Policy 557 [67]. All Facebook data analyzed in this study were obtained from public Facebook pages and groups only. No private or identifiable user information was collected, and data were analyzed in aggregate and were fully deidentified prior to analysis, ensuring the protection of user privacy and compliance with ethical research standards for the secondary use of data.

Data Filtering

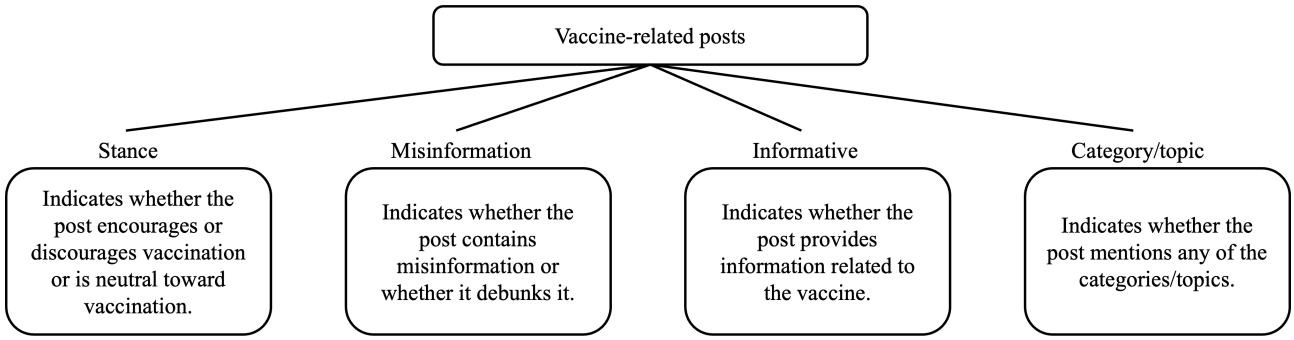
Although querying based on vaccine-related keywords yielded many Facebook posts, an essential step before further data processing involved ensuring that the posts obtained were related to COVID-19 vaccines. Initially, a random sample of 1000 English posts and 500 Spanish posts was extracted from the whole dataset. Then, 2 pairs of annotators (2 annotators for the English posts and 2 for the Spanish posts) manually labeled the samples by indicating whether each post was related to COVID-19 vaccines. The annotators labeling the English dataset were English speakers, with one annotator having a computer science background and the other having a public health background. The annotators labeling the Spanish dataset were Spanish and English speakers, with one annotator having a computer science background and the other having a public health background. The annotators provided a simple binary label for each post, with "1" used to indicate a relation with COVID-19 vaccines and "0" used to indicate no relation. Once the labeling was completed, the annotators met to discuss and resolve any disagreements in labeling. The Cohen κ scores for the labeling of the English and Spanish datasets were 0.76 (substantial agreement) and 0.14 (slight agreement), respectively. The disagreements in the Spanish sample were manually checked, and the low Cohen κ was found to be mainly due to one of the annotators suggesting that posts of job postings mentioning vaccination requirements in the application process should be considered vaccine-related. However, the other annotators did not agree, as these posts did not provide any further discussion regarding the vaccines. Hence, the annotators decided that these posts should not be classified as vaccine-related. In the English dataset, 60.5% (605/1000) of posts were related to COVID-19 vaccines, while in the Spanish dataset, 83.2% (416/500) of posts were related to the vaccines. By manually examining non–vaccine-related posts, lists of keywords relevant to such posts were created.

Some manually identified keywords specific to non–vaccine-related posts were *dog*, *heartworm*, *hiring*, *music*, etc, corresponding to posts related to animals, events that required COVID-19 vaccination, job openings, or document services. After stemming the texts of the posts and the keywords, posts containing any of the keywords from the corresponding lists were classified as non–vaccine-related. Using this methodology, the percentages of misclassified posts when comparing the obtained labels to the ground-truth dataset were 5.9% (59/1000) and 5.6% (28/500) for the English and Spanish datasets, respectively. Therefore, we filtered the whole dataset using these keywords (ie, discarded posts irrelevant to COVID-19 vaccination). We eventually classified 13,819 English posts and 1249 Spanish posts as relevant.

Post Classification Ground-Truth Dataset

Previous literature has suggested many reasons for vaccine hesitancy, such as the fear of potential vaccine side effects, efficacy issues, mistrust toward the government or health system, misinformation spread, and others [11,14,26,68]. Thus, this study assessed English and Spanish Facebook posts to investigate how stance, misinformation, informative versus noninformative posts, and post categories/topics (such as vaccine side effects, government, health system, and others) were associated with vaccination rates in Tarrant County in 2021 and 2022. In more detail, the goal was to identify the potential reasons behind vaccine hesitancy in social media posts and examine the association between online activity and actual vaccination uptake. A summary of relevant variables related to social media posts is illustrated in Figure 2.

Figure 2. Description of the variables used to annotate the posts.



This study attempted to cover all the reasons for vaccine hesitancy and label them in Facebook posts. However, it is important to note that direct measurement of vaccine trust is challenging, despite its significant role in vaccine reluctance [11]. Therefore, the most prevalent reasons for vaccine hesitancy were incorporated as topics in this study.

Additional classifiers were required to annotate each post with the correct labels for each variable. Initially, random samples of 200 posts were extracted from both the English and Spanish datasets. Then, 2 English speakers labeled the English sample (one with a computer science background and the other with a background in statistics), while 2 Spanish speakers (with public health and computer science backgrounds) and 1 English speaker (with a computer science background), who translated the posts, labeled the Spanish sample using the variables described in Figure 2. The labeling codebook is provided in Multimedia Appendix 1. It was used to train the annotators in performing the labeling. The codebook was initially created by incorporating the knowledge from previous literature, which also helped in

XSL•FO
RenderX

defining variables in the Variables section. Manual labeling and defining the codebook were iterative processes. Definitions had to be adjusted multiple times to ensure accuracy and interannotator consistency, and some categories were added by open coding, as discussed in the Independent Variables section. The following are 2 examples from the codebook, including variable definitions and, where applicable, representative examples or supplementary materials:

- Side effects: if the post mentions specific vaccine-related side effects (fever, chills, vomiting, etc; see more here: [69])
- Vaccine availability: if the post mentions anything related to whether vaccinations are available, and where, when, and how. An example post would be "There is a shortage of doses of Pfizer vaccine in Tarrant County."

If the post was irrelevant to COVID-19 vaccination (as some posts remained after keyword filtering), the annotators did not annotate the post. In some instances, annotators had to open the post. They read the title of the article shared within the post, as sometimes the text of the post did not explicitly mention the vaccine, but it still commented on the article about it. Such posts were considered relevant. After the labeling was completed, the annotators met to resolve any conflicts that were raised in the labeling process to create final labels. However, as 3 annotators labeled the Spanish dataset, the final values were computed by a majority vote.

As the goal of this study was to provide an understanding of the differences in COVID-19 vaccine hesitancy between the English and Spanish datasets, the differences in the observed categories found in these datasets have been further discussed. The percentage of posts in each variable category is reported in Table 1. Moreover, examples of posts in each category can be found in Multimedia Appendix 2.

Table . Posts containing each of the relevant labels in the manually labeled sample.

Variable	English (n=185), n (%)	Spanish (n=189), n (%)
Stance		
Encouraging	41 (22.2)	14 (7.4)
Discouraging	6 (3.2)	1 (0.5)
Neither	138 (74.6)	174 (92.1)
Category		
Side effects	3 (1.6)	10 (5.3)
Vaccine availability	89 (48.1)	68 (35.9)
Vaccine safety	7 (3.8)	15 (7.9)
Vaccine benefits/efficacy	22 (11.9)	15 (7.9)
COVID-19 illness experience	6 (3.2)	5 (2.6)
Government	32 (17.3)	23 (12.2)
Education	18 (9.7)	5 (2.6)
Health system	52 (28.1)	20 (10.1)
Religion	3 (1.6)	1 (0.5)
Postvaccination advice	4 (2.2)	5 (2.6)
Community-specific advice	8 (4.3)	13 (6.9)
Policies/mandates	50 (27.0)	47 (24.9)
Statistics	23 (12.4)	8 (4.2)
Informative	150 (81.1)	166 (89.2)
Misinformation		
Contains misinformation	0 (0)	0 (0)
Debunks misinformation	4 (2.2)	1 (0.5)

Interestingly, statistics suggested a substantially greater number of posts in the English manually labeled dataset than the Spanish dataset that either encouraged (41/185, 22.2% vs 14/189, 7.4%) or discouraged (6/185, 3.2% vs 1/189, 0.5%) vaccination. This indicates that the majority of posts in the Spanish manually annotated ground-truth dataset were neutral in their stance toward vaccination, which is also implied by the larger percentage of informative posts in this dataset. Furthermore, a larger percentage of Spanish posts discussed vaccine side effects and safety compared with English posts, which tended to discuss matters related to the educational system, government, policy, and health system. This potentially indicates that the Spanish dataset included posts that mainly focused on the vaccine and its relevant information, while the English dataset included

XSL•FO
RenderX

many posts relevant to the vaccine but also had posts on other aspects connected to its administration. An additional analysis is required to understand how these relevant categories have been discussed over time and what their associations are with vaccination rates in Tarrant County.

ChatGPT Labeling

Once the ground-truth dataset was created, the next step involved obtaining labels for the entire dataset. Large language models (LLMs) have been widely used in the natural language processing literature for text classification tasks, while ChatGPT (OpenAI) has been used for data annotation [70-73]. Moreover, GPT-4 has been shown to be a great choice for detecting COVID-19 vaccine self-report and vaccine chatter [74], reaching accuracies of approximately 81% and 93%, respectively, without any additional prompting or fine-tuning. Therefore, it was a suitable candidate for annotating the vaccine-related dataset, considering its powerful ability to understand the context of the provided input and generate correct labels. Thus, the first step involved formulating a prompt that provided a detailed definition of each variable and corresponding examples, which would enable the model to accurately classify each post. Prompt formulation required several iterations, and the initial prompts were modified so that GPT-4 (OpenAI Azure endpoint was used) could better understand the task and generate labels that aligned more with the ground-truth dataset. By adopting a chain-of-thought [75] approach, larger prompts were broken down into multiple smaller, focused prompts, enabling the model to process each variable in steps. Hence, the labeling codebook was used as a starting point, and the same human thought process was followed to create the prompts that could accurately label the data. In more detail, each prompt included the definition of the variable, an example post belonging to the positive class, and a task to classify the post as related to that variable or not, without explanation. In some cases, the prompt had to explicitly tell the model which posts should not fall under the positive class. For example, the following prompt asked GPT-4 to classify a post as policy/mandate-related or not (the post included a definition of the variable, examples of posts that should be labeled as belonging to the positive class, and exceptions, ie, posts that were previously labeled as belonging to the positive class but should belong to the negative class):

> *Label the text with "Policy related" or "Not policy related." "Policy related" should be used if the post mentions vaccination mandates and requirements, authorizations, recommendations, or other COVID-19 policies/mandates issued by health governmental agencies like the CDC or FDA. This includes employee vaccination requirements, vaccine schedule, travel vaccination mandates, mask mandates, vaccine eligibility, vaccine authorizations, social distancing policies, or public health recommendations such as CDC guidelines for vaccination. Posts that primarily mention where vaccines are available for administration or the logistics of vaccine distribution without discussing mandates, requirements, or authorizations should be labeled as "Not policy related." Respond only with the labels "Policy related" or "Not policy related." Here is the post:*

This example illustrates the need to provide a very specific prompt to the model to obtain annotations that closely match human labels and do not alternate over multiple iterations. Thus, it was necessary to present the model with a simple task (ie, to provide binary labels) and explain the reasoning aspects that should lead the model to the correct class inference. This sequential breakdown improved accuracy (from 81% to 90% in the case of the policy/mandate variable), as the model could more effectively reason through each part of the task to produce consistent labels. The list of final prompts used to obtain labels for each post in the ground-truth dataset can be found in Multimedia Appendix 3. It is important to acknowledge that this annotation task was not simple even for human annotators, and it required multiple discussions and changes in the codebook. This shows the importance of providing very detailed and precise guidelines to both human annotators and LLMs to obtain highly accurate labels.

Once again, GPT-4 was asked to classify posts as "related" or "not related" to the COVID-19 vaccine as an additional filtering step to remove any noise remaining after the initial keyword filtering. In addition to the post text, the titles of the websites shared in the posts were passed to the prompt along with article/site descriptions to provide additional context that annotators could see during the labeling process. This approach was a more accurate method to detect vaccine-related posts, as in some instances, the post text did not directly mention the COVID-19 vaccine but rather commented on an article about the vaccine. Therefore, such posts were also labeled by annotators as vaccine-related. By providing these inputs, GPT-4 could detect such posts with higher accuracy. The accuracies for detecting vaccine-related and non–vaccine-related posts in the English and Spanish datasets were 99% and 96%, respectively. Before calculating the accuracies for other variables, detected non–vaccine-related posts were removed from the English and Spanish ground-truth datasets, and they remained with 185 and 189 posts, respectively. In addition, some posts that were not accurately detected by the classifier as non–vaccine-related (1 post in the English dataset and 3 posts in the Spanish dataset) were also removed when computing the final model accuracy for the rest of the variables, as such posts were not annotated in the manual labeling process. Moreover, the process of obtaining GPT-4 labels for all variables was repeated thrice to ensure that the labels were consistent, as the "hallucination" of LLMs (generating plausible yet nonfactual content) is a known problem in the literature [76]. The majority vote of 3 labels was used to compute the final labels, as the goal was to check how consistent the output was across multiple iterations. The highest percentages of posts in the English dataset that received different labels in these iterations were found to belong to the vaccine availability (16/200, 8.0%) and government (13/200, 6.5%) categories, while for all the other labels, the value did not exceed 6%, with the lowest being around 1% for COVID-19 illness experience, side effects, and religion. On the other hand, for all the variables in the Spanish dataset, the value did not exceed 6%. When obtaining the labels for the entire datasets, the labels were obtained only once, as a limited number of posts received different labels in multiple iterations.

XSL•FO
RenderX

Model Evaluation

After the final labels were computed, they were compared with the manually labeled sample. As demonstrated in Table 2, GPT-4 performed very well in labeling the data with the discussed variables. In the English dataset, the highest accuracy of 99% was obtained for detecting vaccine side effect posts, while the lowest accuracy of 89% was obtained for detecting health system–related posts. However, the classification accuracies in the Spanish dataset were slightly different. In some instances, GPT-4 performed better in the Spanish dataset than in the English dataset (eg, education and statistics), but it did not perform as well in detecting vaccine availability. A possible reason for these results is that the Spanish dataset did not contain as many categories as the English dataset. In some cases, this made the categories harder to detect, while in certain instances, the model could easily recognize the absence of categories, yielding a higher accuracy. It is important to acknowledge that most state-of-the-art classifiers for different text classification tasks have certain limitations; hence, it is not surprising that this classifier performs differently in detecting different categories. As the study's goal did not involve enhancing the state-of-the-art model in detecting the mentioned variables, the classifier's performance will need to be improved in future work.

Table . Accuracy of GPT-4 labels for detecting relevant classes in both English and Spanish datasets.

Variable	Accuracy in the English dataset, %	Accuracy in the Spanish dataset, %
Stance	91	89
Category		
Side effects	99	98
Vaccine availability	90	85
Vaccine safety	97	94
Vaccine benefits/efficacy	95	95
COVID-19 illness experience	98	97
Government	93	90
Education	96	98
Health system	89	86
Religion	97	98
Postvaccination advice	97	96
Community-specific advice	95	93
Policies/mandates	90	88
Statistics	91	98
Informative	90	86
Misinformation	93	93

After labels were obtained for the entire dataset, a small number of posts did not have the expected output, and the model could not infer the corresponding label. This issue mainly occurred for very short posts or posts only containing URLs. These posts were discarded from the datasets (62 posts from the English dataset and 23 from the Spanish dataset).

In generating labels for the entire dataset, the English dataset involved 35.2 million input tokens (prompts and posts) and 307,000 output tokens, while the Spanish dataset involved 2.3 million input tokens (prompts and posts) and 37,000 output tokens. The estimated cost of obtaining the annotations for the entire English dataset was US $1074.8, while the cost for the Spanish dataset was US $71.5. The rate limit allowed sending 10 requests per minute; thus, 14,400 prompts could be sent a day. Considering that 11 different prompts were used in this study (with approximately 15,000 posts), approximately 12 days were required to obtain all the annotations. Considering that the posts were collected over a 2-year period, the cost could be considered acceptable in emergency scenarios where a better understanding of reluctance toward interventions could potentially prevent a public health crisis. Moreover, GPT-4 was the latest model released at the time of this study. However, other open-source models could be explored if enough data are available, and this remains a promising area for future work.

Additional Manual Verification

As noted in Table 1, the distribution of posts for each variable was highly imbalanced. In more detail, most of the time, posts belonging to a negative class for each variable were prevalent in the dataset, which could have led to the misleading accuracies in Table 2. Thus, after completion of labeling of the entire dataset using GPT-4, a random sample of 25 posts belonging to each positive class and 25 posts belonging to each negative class was extracted to verify the accuracy of the annotations. A single annotator labeled 50 posts for each variable (where each class of a multicategorical variable was treated as a binary variable) and compared the labels with the annotations from GPT-4. The results suggested that for most of the variables, GPT-4 annotated 40 or more posts with the accurate label when,

XSL•FO
RenderX

in certain cases, GPT-4 labeled all 50 posts correctly (COVID-19 illness experience and religion in the Spanish dataset). In the Spanish dataset, GPT-4 had low accuracy in detecting misinformation (0.70), postvaccination advice (0.60), and discouraging, informative, and community-specific advice (0.76). In the English dataset, the model had high accuracy in detecting religion (0.98), debunking (0.96), statistics (0.96), and vaccine safety (0.94), and had low accuracy in detecting postvaccination advice (0.62), misinformation (0.74), encouraging advice (0.78), and community-specific advice (0.80). For the remaining variables, the model had accuracies higher than 0.80. Therefore, future work should aim to improve the performance of the classifier for these specific annotations. While the use of GPT-4 for annotating the dataset is highly innovative, the lower accuracies observed for some categories suggest room for improvement, which can be addressed in future work. The performances reported in this study are similar to those reported in related work [74], where for detecting COVID-19 vaccine chatter and self-reported vaccination using GPT-4, the accuracies were nearly 93% and 81%, respectively.

Furthermore, another recent study reported that GPT-4 had an average accuracy of 79.2% in various manually labeled datasets [70]. Hence, it is evident that the model performs better for certain variables than others, where fine-tuning might be required. Further labeling of high-quality datasets and ensuring consistency in the labeling process through prompt tuning might be needed to enhance model performance.

Descriptive Statistics

Overview

This section details the number of posts belonging to each category over the time frame of the data collection. The goal was to compare the prevalence of each post category between 2021 and 2022, and between English and Spanish. Finally, combining some of the categories could provide additional insights into the reasons for vaccine hesitancy and how these have changed over time. It is important to note that the posts classified as unrelated by the model were discarded before the analysis. The cleaned English and Spanish datasets consisted of 12,395 and 1123 posts, respectively.

English Dataset

The majority of posts in the English dataset were from 2021. In more detail, of the 12,395 posts, 11,088 (89.5%) were shared in 2021 and 1307 (10.5%) were shared in 2022. This decline in 2022 may be attributed to the initial surge in vaccine discussions during the roll-out phase in 2021, when uncertainty and debate were more prominent. By 2022, vaccines were widely available, and public attention may have shifted to other topics. The proportion of posts encouraging vaccination was higher in 2021 than in 2022 (2005/11,088, 18.1% vs 154/1307, 11.8%), while the proportion of discouraging posts was lower in 2021 than in 2022 (624/11,088, 5.6% vs 113/1307, 8.6%). This might indicate that at the beginning of vaccine administration in 2021, there was a greater need to encourage vaccination, and thus, a higher percentage of the population decided to receive the vaccine. On the other hand, in 2022, many people had already received the vaccine, and there was no need for a large encouragement. Similarly, the proportion of informative posts was higher in 2021 than in 2022 (8482/11,088, 76.5% vs 943/1307, 72.1%), which might support the earlier assumption. Finally, the proportion of posts containing misinformation was lower in 2021 than in 2022 (414/11,088, 3.7% vs 79/1307, 6.0%).

Table 3 presents the proportion of encouraging and discouraging posts for each category in 2021 and 2022 in the English dataset. As demonstrated in Table 3, the categories of encouraging and discouraging posts were very different. For instance, the majority of encouraging posts in both 2021 and 2022 were informative, and they mostly discussed vaccine availability and vaccine benefits/efficacy. These posts also promoted the health system, offered advice to specific communities and minorities, and provided useful statistics to encourage vaccination. The topic of education was more common among encouraging posts, which might be due to the high support provided for student vaccination to allow them to continue regular studying activities and in-person classes. On the other hand, discouraging posts in both 2021 and 2022 mostly discussed the government and policies/mandates regarding vaccination, suggesting that the population was hesitant toward the measures taken to mitigate the pandemic, such as vaccination requirements, vaccination passports, and mask mandates. A large proportion of discouraging posts involved misinformation in 2021 (273/624, 43.8%) and 2022 (57/113, 50.4%). A slightly lower proportion of discouraging posts discussed the health system and vaccine safety, and they rarely offered advice or debunked misinformation. Despite these proportions being relatively low, the proportion of posts regarding religion and side effects was higher among discouraging posts than among encouraging posts, suggesting that these factors might still play significant roles in vaccine hesitancy.

The data suggested that misinformation and concerns about government policies were major contributors to vaccine hesitancy, while informative posts and discussions about vaccine availability played critical roles in promoting vaccination.

XSL•FO
RenderX

Table . Proportion of posts containing each relevant category among encouraging and discouraging posts in 2021 and 2022 in the English dataset.

Variable	2021		2022	
	Encouraging posts (n=2005), n (%)	Discouraging posts (n=624), n (%)	Encouraging posts (n=154), n (%)	Discouraging posts (n=113), n (%)
Side effects	40 (2.0)	36 (5.8)	1 (0.6)	5 (4.4)
Vaccine availability	809 (40.3)	16 (2.6)	77 (50.0)	4 (3.5)
Vaccine safety	130 (6.5)	73 (11.7)	2 (1.3)	13 (11.5)
Vaccine benefits/efficacy	740 (36.9)	26 (4.2)	77 (50.0)	3 (2.7)
COVID-19 illness experience	49 (2.3)	2 (0.3)	1 (0.6)	0 (0)
Government	150 (7.5)	287 (46.0)	14 (9.1)	59 (52.2)
Education	116 (5.8)	25 (4.0)	14 (9.1)	1 (0.8)
Health system	714 (35.6)	132 (21.2)	61 (39.6)	28 (24.8)
Religion	49 (2.4)	32 (5.1)	1 (0.6)	3 (2.7)
Vaccination advice	156 (7.8)	4 (0.6)	10 (6.5)	0 (0)
Community-specific advice	39 (1.9)	10 (1.6)	3 (1.9)	0 (0)
Policies/mandates	361 (18.0)	334 (53.5)	38 (24.7)	70 (61.9)
Statistics	184 (9.2)	18 (2.9)	17 (11.0)	2 (1.8)
Informative	1766 (88.1)	209 (33.5)	148 (96.1)	33 (29.2)
Contains misinformation	10 (0.5)	273 (43.8)	1 (0.6)	57 (50.4)
Debunks misinformation	32 (1.5)	9 (1.4)	1 (0.6)	0 (0)

Spanish Dataset

The total number of posts was lower in the Spanish dataset than in the English dataset. Of the 1123 posts, 984 (87.6%) were shared in 2021 and 139 (12.4%) were shared in 2022. The proportion of posts encouraging vaccination was higher in 2021 than in 2022 (158/984, 16.1% vs 14/139, 10.1%), while the proportion of discouraging posts was lower in 2021 than in 2022 (58/984, 5.9% vs 12/139, 8.6%). The proportion of informative posts was higher in 2021 than in 2022 (830/984, 84.3% vs 111/139, 79.9%), while the proportion of posts containing misinformation was lower in 2021 than in 2022 (44/984, 4.5% vs 14/139, 10.1%).

As illustrated in Table 4, encouraging posts in both 2021 and 2022 discussed vaccine availability, benefits/efficacy, the government, and policies, and offered community-specific advice. Interestingly, the topic of the government was more prevalent among encouraging posts than among discouraging posts in both 2021 and 2022, potentially suggesting that mistrust toward the government was not a major cause of vaccine hesitancy in the Hispanic population in Tarrant County. A larger proportion of discouraging posts mentioned policies/mandates compared with encouraging posts. Moreover, a large proportion of discouraging posts discussed vaccine side effects and vaccine safety, revealing additional potential reasons for vaccine hesitancy, which differ from the statistics found in the English dataset, where discouraging posts focused on the government and mandates. Given the smaller dataset, these findings should be interpreted with caution. However, they reveal that vaccine safety and side effects might be significant concerns for the Hispanic population, distinct from the focus on government policies in the English dataset. The health system was a common topic among both encouraging and discouraging Spanish posts; thus, an additional analysis is required to understand the actual relationship between vaccine hesitancy and this topic. Similar to the English dataset, most discouraging posts in the Spanish dataset contained some misinformation. These descriptive insights set the stage for a statistical analysis, which could further explore the relationship between these categories and vaccination rates.

XSL•FO
RenderX

Table . Proportion of posts containing each relevant category among encouraging and discouraging posts in 2021 and 2022 in the Spanish dataset.

Variable	2021		2022	
	Encouraging posts (n=158), n (%)	Discouraging posts (n=58), n (%)	Encouraging posts (n=14), n (%)	Discouraging posts (n=12), n (%)
Side effects	6 (3.8)	11 (19.0)	0 (0)	2 (16.7)
Vaccine availability	67 (42.4)	1 (1.7)	6 (42.9)	0 (0)
Vaccine safety	8 (5.1)	14 (24.1)	0 (0)	1 (8.3)
Vaccine benefits/efficacy	32 (20.3)	2 (3.4)	3 (21.4)	0 (0)
COVID-19 illness experience	1 (0.6)	2 (3.4)	0 (0)	0 (0)
Government	17 (10.8)	5 (8.6)	2 (14.3)	1 (8.3)
Education	11 (7.0)	1 (1.8)	1 (7.1)	0 (0)
Health system	41 (26.0)	22 (37.9)	5 (35.7)	2 (16.7)
Religion	5 (3.2)	2 (3.4)	0 (0)	0 (0)
Vaccination advice	3 (1.9)	0 (0)	0 (0)	0 (0)
Community-specific advice	16 (10.1)	0 (0)	1 (7.1)	0 (0)
Policies/mandates	23 (14.6)	17 (29.3)	4 (28.6)	4 (33.3)
Statistics	6 (3.8)	4 (6.9)	0 (0)	0 (0)
Informative	143 (90.1)	33 (56.9)	14 (100)	5 (41.7)
Contains misinformation	1 (0.6)	26 (44.8)	0 (0)	8 (66.7)
Debunks misinformation	3 (1.9)	0 (0)	0 (0)	0 (0)

Statistical Analysis

This study aimed to explore how different characteristics of social media posts are associated with vaccination rates in Tarrant County and compare the trends observed in total vaccination rates and Hispanic vaccination rates. In more detail, the analysis focused on the relationship between different social media post features in English and Spanish and vaccination rates at the daily level throughout 2021 and 2022.

For the analysis, English and Spanish posts were merged into a single dataset (containing a total of 13,518 data points), as there might be Hispanic individuals who post in English rather than Spanish. Merging the English and Spanish datasets required careful attention to both linguistic and cultural nuances in vaccine-related discourse. While the same annotation framework and prompts were applied to both datasets using GPT-4, the way topics are framed can differ significantly across languages and communities. For example, the expression of vaccine skepticism in Spanish posts might often include religious or familial references that could be less common in English posts, requiring more context-aware labeling. Additionally, certain categories, such as government trust or perception of the health care system, may carry different connotations based on cultural background, which could influence how posts are interpreted and categorized. To mitigate these challenges, bilingual annotators reviewed posts and labels in both languages to account for this contextual variation. Nevertheless, we acknowledge that achieving complete semantic and cultural equivalence is challenging, and residual bias in cross-language classification remains a limitation. The dataset was then split into posts originating in 2021 and 2022. For each year, 2 linear regression models corresponding to 2 dependent variables were used by clustering SEs per day, as posts shared on the same day might not be independent of each other. Linear regression was selected for this analysis because the outcome variables (ie, daily total vaccinations and daily Hispanic vaccinations) are continuous, allowing us to estimate the linear relationship between social media content features and vaccination uptake. Since each post is assigned the same vaccination count for its corresponding day, multiple posts share the same value for the dependent variable. This creates “within-day clustering,” where residuals of observations from the same day may be correlated, violating the independence assumption of standard linear regression. To address this, we used “clustered SEs at the day level,” which provide robust inference by correcting for intracluster correlation and ensuring more accurate CIs and *P* values. This approach allows for a valid statistical inference even when the number of posts varies substantially across days.

Furthermore, each data point represented a Facebook post and included the binary categorical variables describing the post. Other variables included in each observation were the date when the post was shared, *new Hispanic vaccinations*, *new total vaccinations* on that particular day, and the language of the post. All 4 models included all the independent and corresponding control variables.

XSL•FO
RenderX

Results

Regression Analysis of Social Media Content and Vaccination Outcomes

This section presents the findings of the 4 linear regression models where SEs were clustered per day. It is important to note that similar models were leveraged, but they used the weekly vaccination data rather than the daily data, and SEs were clustered per week. The results of this analysis closely resemble the findings discussed in this section, showing the robustness of the models. Moreover, the models were tested by using a dependent variable, where new vaccinations per day and week were divided by the population available to vaccinate, and the results remained similar. Table 5 presents the significant results of the 4 models in the analysis (variables with nonsignificant findings have been excluded for readability). Complete data are provided in Multimedia Appendix 4.

Table . Results of the linear regression models.

Variable[a]	Estimate	SE[b]	*P* value
Model 1[c] (dependent variable: *new total vaccinations* in 2021)			
Vaccine safety - true	4.00×10^{-4}	1.46×10^{-4}	.006
Government - true	-1.44×10^{-4}	6.81×10^{-5}	.04
Vaccine availability - true	1.58×10^{-4}	6.19×10^{-5}	.01
Population available	3.10×10^{-3}	5.00×10^{-4}	<.001
Model 2[d] (dependent variable: *new Hispanic vaccinations* in 2021)			
Encouraging - true	9.88×10^{-5}	4.25×10^{-5}	.02
Vaccine safety - true	5.27×10^{-4}	1.70×10^{-4}	.002
Government - true	-1.80×10^{-4}	5.37×10^{-5}	<.001
Religion - true	2.42×10^{-4}	8.91×10^{-5}	.007
Vaccine availability - true	1.01×10^{-4}	4.89×10^{-5}	.04
Hispanic population available	1.03×10^{-3}	4.40×10^{-4}	.02
Model 3[e] (dependent variable: *new total vaccinations* in 2022)			
Statistics - true	-5.69×10^{-5}	1.69×10^{-5}	<.001
Postvaccination advice - true	7.27×10^{-5}	3.64×10^{-5}	.046
Debunking - true	2.01×10^{-4}	5.36×10^{-5}	<.001
Population available	1.84×10^{-2}	1.36×10^{-3}	<.001
Model 4[f] (dependent variable: *new Hispanic vaccinations* in 2022)			
Statistics - true	-6.98×10^{-5}	2.09×10^{-5}	<.001
Debunking - true	1.94×10^{-4}	9.56×10^{-5}	.04
Hispanic population available	1.97×10^{-2}	1.60×10^{-3}	<.001

[a]Models 1 and 2 used data from 2021, while models 3 and 4 used data from 2022.
[b]SEs were clustered per day.
[c]R^2=0.13.
[d]R^2=0.04.
[e]R^2=0.70.
[f]R^2=0.68.

Trends in 2021

Regression analysis yielded multiple relevant findings. As demonstrated for models 1 and 2 in Table 5, posts discussing vaccine safety were associated with a higher daily increase in total vaccination (*P*=.006) and vaccination of the Hispanic population (*P*=.002). This might indicate that demonstration of vaccine safety once vaccine administration begins is crucial for ensuring a successful intervention. Furthermore, posts mentioning the government were associated with a lower daily increase in both total (*P*=.04) and Hispanic vaccinations (*P*<.001), potentially suggesting mistrust toward the government

XSL•FO
RenderX

at that time. In addition, posts providing details about vaccine availability increased the likelihood of a higher number of newly vaccinated individuals (*P*=.01) and Hispanic individuals (*P*=.04). Interestingly, posts about religion showed a statistically significant positive relationship with *new Hispanic vaccinations* (*P*=.007), indicating that such posts might increase the probability of new Hispanic individuals obtaining the vaccine. It is important to note that when running the models where vaccination increase was calculated per week, all variables that showed statistical significance at the daily level remained significant. However, posts regarding the health system showed a negative statistically significant correlation with weekly *new total vaccinations* (*P*=.04) and *new Hispanic vaccinations* (*P*=.03), suggesting potential mistrust toward the health system in this period. Additionally, posts shared in Spanish showed a positive statistically significant correlation with weekly *new Hispanic vaccinations* (*P*=.02), illustrating that posts shared in Spanish were associated with a higher vaccination increase in the Hispanic population compared with posts shared in English.

Trends in 2022

Models 3 and 4 (Table 5) from 2022 had different results compared with the models from 2021. For example, posts debunking misinformation showed a statistically significant relationship with both *new total vaccinations* (*P*<.001) and *new Hispanic vaccinations* (*P*=.04), demonstrating that such posts are associated with a higher number of new vaccinations compared with nondebunking posts. It is important to note that the difference in findings could be due to the size of the dataset in 2022 compared with the dataset in 2021. In summary, the results (Table 5) indicate that the stance of social media posts, posts debunking misinformation, and some other informational categories, such as vaccine availability and safety, play significant roles in influencing the number of newly vaccinated individuals.

Discussion

Principal Findings

This study examined bilingual Facebook discourse about COVID-19 vaccines in Tarrant County and linked it to daily vaccination uptake in the total and Hispanic populations. Three main findings stand out.

First, vaccine-related discourse differed substantially by language and year. English and Spanish posts did not emphasize the same concerns, and the mix of topics shifted from 2021 to 2022. English discouraging posts were dominated by government and policy concerns, whereas Spanish discouraging posts more often emphasized side effects, safety, and the health system. This suggests that vaccine hesitancy was not uniform across language groups and that communication strategies should not assume the same drivers across populations.

Second, some predictors of uptake were common across groups. Posts discussing vaccine availability and vaccine safety were positively associated with vaccination in 2021, and posts debunking misinformation were positively associated with vaccination in 2022. These findings are consistent with the view that practical access information, trust-building content, and correction of false claims are important components of effective vaccine communication.

Third, some predictors were specific to Hispanic vaccination uptake. Encouraging posts and religion-related posts were positively associated with *new Hispanic vaccinations* in 2021, but not with total vaccinations in the same way. This is particularly important because it suggests that culturally grounded content may matter above and beyond general informational messaging.

The significance of religion-related posts is consistent with prior literature showing that religiosity can shape health behavior [57]. In some communities, faith leaders and faith-based institutions serve as highly trusted intermediaries. In such settings, messages framed around protecting family, stewardship, and community responsibility may resonate more strongly than messages framed only in technical or institutional terms. These findings suggest that partnerships with churches and faith-based organizations may be especially valuable in public health outreach to Hispanic communities.

The findings of this study provide several actionable recommendations for public health officials. Social media interventions targeting vaccine-hesitant populations, particularly in the Hispanic community, should be designed with cultural sensitivity in mind [58]. For example, incorporating bilingual content and collaborating with trusted community figures—such as religious leaders or local health care providers—could enhance the credibility and reach of these interventions. Furthermore, posts that emphasize family protection, community well-being, and the dispelling of myths related to vaccine safety and efficacy have been shown to resonate well with Hispanic audiences [77]. By focusing on culturally appropriate messaging and leveraging trusted networks within these communities, public health efforts can more effectively counter vaccine misinformation and build confidence in the vaccination process.

In summary, the results underscore the importance of bilingual communication. Because language may shape how health information is understood and trusted, simply translating English-language messages may not be sufficient. Effective outreach may require culturally aligned framing, not just linguistic conversion.

Limitations

This study has several limitations. First, the dataset included only posts from public Facebook pages and groups, which may not represent the views of people using private groups, other platforms, or no social media at all. Second, language is not a perfect proxy for ethnicity, and thus, it cannot be assumed that Spanish-language posts come only from Hispanic users or English-language posts come only from non-Hispanic users. Third, because multiple posts shared the same daily vaccination outcome, the observational structure limits causal interpretation even though clustered SEs were used. Fourth, GPT-4 annotation, while strong overall, was not perfect and may have introduced measurement errors in some categories. Finally, the models did not fully capture broader cultural, structural, or offline influences on vaccine behavior.

XSL•FO
RenderX

Conclusions

This study examined how stance, misinformation, and topic content in English and Spanish Facebook posts were associated with new daily vaccinations in Tarrant County, Texas, during 2021 and 2022. The findings showed that social media discourse differed meaningfully across languages and over time and that some message types were associated with higher vaccination uptake. Across both populations, posts addressing vaccine availability, safety, and misinformation debunking were associated with higher uptake. For the Hispanic population specifically, encouraging posts and religion-related posts were also significant predictors, suggesting culturally specific dynamics.

Taken together, these findings support the use of targeted, culturally responsive, and bilingual public health messaging in future emergencies, and they suggest that social media analysis can help identify which kinds of messages may be most useful for reaching specific communities.

Acknowledgments

We thank Will Chen, Keren Valdez, Breanna Guerrero, and Diana Tostado for their contributions to data annotation. We also thank Henry Isaac Andreson for feedback on data collection and annotation. Generative artificial intelligence was not used to generate any portion of this manuscript.

Funding

This research received partial funding (project number: 1NH75OT000054-01-00) through Tarrant County Public Health/CDC (project title: National Initiative to Address COVID-19 Health Disparities Among Populations at High-Risk and Underserved, Including Racial and Ethnic Minority Populations and Rural Communities).

Data Availability

The vaccination datasets analyzed in this study are not publicly available due to access restrictions imposed by Tarrant County. Facebook data are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Labeling guidelines.
[DOCX File, 17 KB - ojphi_v18i1e72465_app1.docx]

Multimedia Appendix 2
Example posts for each category.
[DOCX File, 21 KB - ojphi_v18i1e72465_app2.docx]

Multimedia Appendix 3
List of prompts.
[DOCX File, 18 KB - ojphi_v18i1e72465_app3.docx]

Multimedia Appendix 4
Full regression tables.
[DOCX File, 24 KB - ojphi_v18i1e72465_app4.docx]

References

1. Al-Jayyousi GF, Sherbash MAM, Ali LAM, et al. Factors influencing public attitudes towards COVID-19 vaccination: a scoping review informed by the socio-ecological model. Vaccines (Basel) 2021 May 24;9(6):548. [doi: 10.3390/vaccines9060548] [Medline: 34073757]
2. Greenwood B. The contribution of vaccination to global health: past, present and future. Philos Trans R Soc Lond B Biol Sci 2014;369(1645):20130433. [doi: 10.1098/rstb.2013.0433] [Medline: 24821919]
3. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495. [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]
4. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2024-02-15]

XSL•FO
RenderX

5. Dhama K, Sharun K, Tiwari R, et al. COVID-19 vaccine hesitancy - reasons and solutions to achieve a successful global vaccination campaign to tackle the ongoing pandemic. Hum Vaccin Immunother 2021 Oct 3;17(10):3495-3499. [doi: 10.1080/21645515.2021.1926183] [Medline: 34191680]
6. Sallam M. COVID-19 vaccine hesitancy worldwide: a concise systematic review of vaccine acceptance rates. Vaccines (Basel) 2021 Feb 16;9(2):160. [doi: 10.3390/vaccines9020160] [Medline: 33669441]
7. Allington D, McAndrew S, Moxham-Hall V, Duffy B. Coronavirus conspiracy suspicions, general vaccine attitudes, trust and coronavirus information source as predictors of vaccine hesitancy among UK residents during the COVID-19 pandemic. Psychol Med 2023 Jan;53(1):236-247. [doi: 10.1017/S0033291721001434] [Medline: 33843509]
8. Robinson E, Jones A, Lesser I, Daly M. International estimates of intended uptake and refusal of COVID-19 vaccines: a rapid systematic review and meta-analysis of large nationally representative samples. Vaccine (Auckl) 2021 Apr 8;39(15):2024-2034. [doi: 10.1016/j.vaccine.2021.02.005] [Medline: 33722411]
9. Adu P, Poopola T, Medvedev ON, et al. Implications for COVID-19 vaccine uptake: a systematic review. J Infect Public Health 2023 Mar;16(3):441-466. [doi: 10.1016/j.jiph.2023.01.020] [Medline: 36738689]
10. Nguyen KH, Srivastav A, Razzaghi H, et al. COVID-19 vaccination intent, perceptions, and reasons for not vaccinating among groups prioritized for early vaccination - United States, September and December 2020. Am J Transplant 2021 Apr;21(4):1650-1656. [doi: 10.1111/ajt.16560] [Medline: 33788992]
11. Willis DE, Andersen JA, Bryant‐Moore K, et al. COVID‐19 vaccine hesitancy: race/ethnicity, trust, and fear. Clinical Translational Sci 2021 Nov;14(6):2200-2207. [doi: 10.1111/cts.13077]
12. Matthews KRW, Lakshmanan R. Vaccine hesitancy and resistance in Texas: an analysis of testimony from the 2021 Texas state legislative session. : Rice University's Baker Institute for Public Policy; 2022. [doi: 10.25613/NE2G-HK67]
13. Artiga S. Growing gaps in COVID-19 vaccinations among Hispanic people. The Kaiser Family Foundation. 2021. URL: https://www.kff.org/policy-watch/growing-gaps-covid-19-vaccinations-hispanic-people/ [accessed 2024-02-20]
14. Rosbach M. Medical mistrust, misinformation factor into Latino vaccine hesitancy, OSU study finds. Oregon State University. 2021. URL: https://today.oregonstate.edu/news/medical-mistrust-misinformation-factor-latino-vaccine-hesitancy-osu-study-finds [accessed 2026-04-10]
15. Jayasurya GG, Kumar S, Singh BK, Kumar V. Analysis of public sentiment on COVID-19 vaccination using Twitter. IEEE Trans Comput Soc Syst 2022;9(4):1101-1111. [doi: 10.1109/TCSS.2021.3122439]
16. Jemielniak D, Krempovych Y. An analysis of AstraZeneca COVID-19 vaccine misinformation and fear mongering on Twitter. Public Health (Fairfax) 2021 Nov;200:4-6. [doi: 10.1016/j.puhe.2021.08.019] [Medline: 34628307]
17. Marcec R, Likic R. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. Postgrad Med J 2022 Jul;98(1161):544-550. [doi: 10.1136/postgradmedj-2021-140685] [Medline: 34373343]
18. Ortiz-Sánchez E, Velando-Soriano A, Pradas-Hernández L, et al. Analysis of the anti-vaccine movement in social networks: a systematic review. Int J Environ Res Public Health 2020 Jul 27;17(15):5394. [doi: 10.3390/ijerph17155394] [Medline: 32727024]
19. Alam KN, Khan MS, Dhruba AR, et al. Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data. Comput Math Methods Med 2021;2021:4321131. [doi: 10.1155/2021/4321131] [Medline: 34899965]
20. Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera J, et al. Vaccine hesitancy on social media: sentiment analysis from June 2011 to April 2019. Vaccines (Basel) 2021 Jan 7;9(1):28. [doi: 10.3390/vaccines9010028] [Medline: 33430428]
21. Ali G, Rahman MM, Hossain MA, et al. Public perceptions of COVID-19 vaccines: policy implications from US spatiotemporal sentiment analytics. Healthcare (Basel) 2021 Aug 27;9(9):1110. [doi: 10.3390/healthcare9091110] [Medline: 34574884]
22. Chen Q, Crooks A. Analyzing the vaccination debate in social media data pre- and post-COVID-19 pandemic. Int J Appl Earth Obs Geoinf 2022 Jun;110:102783. [doi: 10.1016/j.jag.2022.102783] [Medline: 35528967]
23. Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: a call to action for strengthening vaccine confidence. J Infect Public Health 2021 Oct;14(10):1505-1512. [doi: 10.1016/j.jiph.2021.08.010] [Medline: 34426095]
24. 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. [doi: 10.1016/j.ijid.2021.05.059] [Medline: 34052407]
25. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned sentiment analysis of COVID-19 vaccine-related social media data: comparative study. J Med Internet Res 2022 Oct 17;24(10):e40408. [doi: 10.2196/40408] [Medline: 36174192]
26. Muric G, Wu Y, Ferrara E. COVID-19 vaccine hesitancy on social media: building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642. [doi: 10.2196/30642] [Medline: 34653016]
27. Ngai CSB, Singh RG, Yao L. Impact of COVID-19 vaccine misinformation on social media virality: content analysis of message themes and writing strategies. J Med Internet Res 2022 Jul 6;24(7):e37806. [doi: 10.2196/37806] [Medline: 35731969]

XSL•FO
RenderX

28. Liew TM, Lee CS. Examining the utility of social media in COVID-19 vaccination: unsupervised learning of 672,133 Twitter posts. JMIR Public Health Surveill 2021 Nov 3;7(11):e29789. [doi: 10.2196/29789] [Medline: 34583316]
29. Aleksandric A, Obasanya MJ, Melcher S, Nilizadeh S, Wilson GM. Your tweets matter: how social media sentiments associate with COVID-19 vaccination rates in the US. Online J Public Health Inform 2022;14(1):e2. [doi: 10.5210/ojphi.v14i1.12419] [Medline: 36120165]
30. Aleksandric A, Anderson HI, Melcher S, Nilizadeh S, Wilson GM. Spanish Facebook posts as an indicator of COVID-19 vaccine hesitancy in Texas. Vaccines (Basel) 2022 Oct 14;10(10):1713. [doi: 10.3390/vaccines10101713] [Medline: 36298580]
31. Herrera-Peco I, Jiménez-Gómez B, Peña Deudero JJ, Benitez De Gracia E, Ruiz-Núñez C. Healthcare professionals' role in social media public health campaigns: analysis of Spanish pro vaccination campaign on Twitter. Healthcare (Basel) 2021 Jun 2;9(6):662. [doi: 10.3390/healthcare9060662] [Medline: 34199495]
32. Herrera-Peco I, Jiménez-Gómez B, Romero Magdalena CS, et al. Antivaccine movement and COVID-19 negationism: a content analysis of Spanish-written messages on Twitter. Vaccines (Basel) 2021 Jun 15;9(6):656. [doi: 10.3390/vaccines9060656] [Medline: 34203946]
33. Ruiz-Núñez C, Segado-Fernández S, Jiménez-Gómez B, et al. Bots' activity on COVID-19 pro and anti-vaccination networks: analysis of Spanish-written messages on Twitter. Vaccines (Basel) 2022 Aug 2;10(8):1240. [doi: 10.3390/vaccines10081240] [Medline: 36016126]
34. Valdez D, Soto-Vásquez AD, Montenegro MS. Geospatial vaccine misinformation risk on social media: online insights from an English/Spanish natural language processing (NLP) analysis of vaccine-related tweets. Soc Sci Med 2023 Dec;339:116365. [doi: 10.1016/j.socscimed.2023.116365] [Medline: 37984184]
35. Catalan-Matamoros D, Prieto-Sanchez I, Langbecker A. Crisis communication during COVID-19: English, French, Portuguese, and Spanish discourse of AstraZeneca vaccine and Omicron variant on social media. Vaccines (Basel) 2023 Jun 15;11(6):1100. [doi: 10.3390/vaccines11061100] [Medline: 37376489]
36. Schmid-Petri H, Bürger M, Schlögl S, Schwind M, Mitrović J, Kühn R. The multilingual Twitter-discourse on vaccination in Germany during the COVID-19 pandemic. Media Commun 2023;11(1):293-305. [doi: 10.17645/mac.v11i1.6058]
37. Lenti J, Mejova Y, Kalimeri K, et al. Global misinformation spillovers in the vaccination debate before and during the COVID-19 pandemic: multilingual Twitter study. JMIR Infodemiology 2023 May 24;3:e44714. [doi: 10.2196/44714] [Medline: 37223965]
38. Quick facts: Tarrant County, Texas. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/tarrantcountytexas/PST045223 [accessed 2026-04-10]
39. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
40. McCombs ME, Shaw DL. The agenda-setting function of mass media. Public Opin Q 1972;36(2):176. [doi: 10.1086/267990]
41. ALDayel A, Magdy W. Stance detection on social media: state of the art and trends. Inf Process Manag 2021 Jul;58(4):102597. [doi: 10.1016/j.ipm.2021.102597]
42. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine (Auckl) 2017 May 25;35(23):3033-3040. [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
43. Dyda A, Shah Z, Surian D, et al. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. Hum Vaccin Immunother 2019 Aug 3;15(7-8):1488-1495. [doi: 10.1080/21645515.2019.1596712]
44. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Soc Sci Med 2017 Oct;191:168-175. [doi: 10.1016/j.socscimed.2017.08.041]
45. Wu L, Morstatter F, Carley KM, Liu H. Misinformation in social media: definition, manipulation, and detection. ACM SIGKDD Explorations Newsletter 2019;21(2):80-90. [doi: 10.1145/3373464.3373475]
46. Ullah I, Khan KS, Tahir MJ, Ahmed A, Harapan H. Myths and conspiracy theories on vaccines and COVID-19: potential effect on global vaccine refusals. Vacunas 2021;22(2):93-97. [doi: 10.1016/j.vacun.2021.01.001] [Medline: 33727904]
47. Kricorian K, Civen R, Equils O. COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety. Hum Vaccin Immunother 2022 Dec 31;18(1):1950504. [doi: 10.1080/21645515.2021.1950504] [Medline: 34325612]
48. Pierri F, Perry BL, DeVerna MR, et al. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. Sci Rep 2022 Apr 26;12(1):5966. [doi: 10.1038/s41598-022-10070-w] [Medline: 35474313]
49. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med 2020 Oct;263:113356. [doi: 10.1016/j.socscimed.2020.113356]
50. Carrieri V, Guthmuller S, Wübker A. Trust and COVID-19 vaccine hesitancy. Sci Rep 2023 Jun 7;13(1):9245. [doi: 10.1038/s41598-023-35974-z] [Medline: 37286569]
51. Lee SK, Sun J, Jang S, Connelly S. Misinformation of COVID-19 vaccines and vaccine hesitancy. Sci Rep 2022 Aug 11;12(1):13681. [doi: 10.1038/s41598-022-17430-6] [Medline: 35953500]
52. Yang Z, Luo X, Jia H. Is it all a conspiracy? Conspiracy theories and people's attitude to COVID-19 vaccination. Vaccines (Basel) 2021 Sep 22;9(10):1051. [doi: 10.3390/vaccines9101051] [Medline: 34696159]

XSL•FO
RenderX

53. Luo X, Jia H. When scientific literacy meets nationalism: exploring the underlying factors in the Chinese public's belief in COVID-19 conspiracy theories. Chinese Journal of Communication 2022 Apr 3;15(2):227-249. [doi: 10.1080/17544750.2021.1954963]
54. Jarrett C, Wilson R, O'Leary M, Eckersberger E, Larson HJ. Strategies for addressing vaccine hesitancy – a systematic review. Vaccine (Auckl) 2015 Aug;33(34):4180-4190. [doi: 10.1016/j.vaccine.2015.04.040]
55. Peters MDJ. Addressing vaccine hesitancy and resistance for COVID-19 vaccines. Int J Nurs Stud 2022 Jul;131:104241. [doi: 10.1016/j.ijnurstu.2022.104241] [Medline: 35489108]
56. Kalra S, Kalra D, Grafova I, et al. Association of death or illness from COVID-19 among family and friends on vaccine uptake within four months of the emergency use authorization. Findings from a national survey in the United States. Vaccine (Auckl) 2023 Mar 17;41(12):1911-1915. [doi: 10.1016/j.vaccine.2023.01.024] [Medline: 36682984]
57. Garcia LL, Yap JFC. The role of religiosity in COVID-19 vaccine hesitancy. J Public Health (Bangkok) 2021 Sep 22;43(3):e529-e530. [doi: 10.1093/pubmed/fdab192]
58. Smith DG, Smith CD, DeLeon JA, et al. Factors influencing COVID-19 vaccine uptake among Latinos: a cross-sectional study. PLoS ONE 2024;19(7):e0302332. [doi: 10.1371/journal.pone.0302332] [Medline: 38968176]
59. Aleksandric A, Dangal A, Nilizadeh S, Mustata Wilson G. Facebook post credibility as a predictor of vaccine hesitancy in the US. Stud Health Technol Inform 2024 Jan 25;310:981-985. [doi: 10.3233/SHTI231111] [Medline: 38269955]
60. Karabela Ş, Coşkun F, Hoşgör H. Investigation of the relationships between perceived causes of COVID-19, attitudes towards vaccine and level of trust in information sources from the perspective of Infodemic: the case of Turkey. BMC Public Health 2021 Jun 23;21(1):1195. [doi: 10.1186/s12889-021-11262-1] [Medline: 34158015]
61. Mønsted B, Lehmann S. Characterizing polarization in online vaccine discourse-a large-scale study. PLoS ONE 2022;17(2):e0263746. [doi: 10.1371/journal.pone.0263746] [Medline: 35139121]
62. Facebook. URL: https://www.facebook.com/ [accessed 2024-02-21]
63. Social media fact sheet. Pew Research Center. 2025. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2026-04-10]
64. CrowdTangle. GitHub. URL: https://github.com/CrowdTangle/API [accessed 2026-04-20]
65. DeVerna MR, Pierri F, Truong BT, et al. CoVaxxy: a collection of English-language Twitter posts about COVID-19 vaccines. Proceedings of the International AAAI Conference on Web and Social Media 2021;15:992-999. [doi: 10.1609/icwsm.v15i1.18122]
66. U.S. Department of Health and Human Services. 45 CFR 46.104 (d)(2i): Exempt research. Code of Federal Regulations. URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46/subpart-A/section-46.104 [accessed 2026-04-20]
67. Centers for Disease Control and Prevention (U.S.), Office of the Associate Director for Science. Distinguishing public health research and public health nonresearch (Policy 557). Centers for Disease Control and Prevention (US). URL: https://stacks.cdc.gov/view/cdc/24235 [accessed 2026-04-20]
68. Troiano G, Nardi A. Vaccine hesitancy in the era of COVID-19. Public Health (Fairfax) 2021 May;194:245-251. [doi: 10.1016/j.puhe.2021.02.025] [Medline: 33965796]
69. COVID-19 vaccine VIS. Centers for Disease Control and Prevention (US). URL: https://www.cdc.gov/vaccines/hcp/vis/vis-statements/COVID-19.html [accessed 2026-04-20]
70. Aldeen M, Luo J, Lian A, et al. ChatGPT vs. human annotators: a comprehensive analysis of ChatGPT for text annotation. Presented at: 2023 International Conference on Machine Learning and Applications (ICMLA); Dec 15-17, 2023. [doi: 10.1109/ICMLA58977.2023.00089]
71. Gilardi F, Alizadeh M, Kubli M. ChatGPT outperforms crowd workers for text-annotation tasks. Proc Natl Acad Sci U S A 2023 Jul 25;120(30):e2305016120. [doi: 10.1073/pnas.2305016120] [Medline: 37463210]
72. Zhu Y, Zhang P, Haq EU, Hui P, Tyson G. Can ChatGPT reproduce human-generated labels? A study of social computing tasks. arXiv. Preprint posted online on Apr 22, 2023 URL: https://arxiv.org/abs/2304.10145 [accessed 2024-11-01]
73. Zhu Y, Yin Z, Tyson G, Haq EU, Lee LH, Hui P. APT-pipe: a prompt-tuning tool for social data annotation using ChatGPT. Presented at: ACM Web Conference 2024; May 13-17, 2024. [doi: 10.1145/3589334.3645642]
74. Tekumalla R, Banda JM. Leveraging large language models and weak supervision for social media data annotation: an evaluation using COVID-19 self-reported vaccination tweets. In: Mori H, Asahi Y, Coman A, Vasilache S, Rauterberg M, editors. HCI International 2023 – Late Breaking Papers HCII 2023 Lecture Notes in Computer Science: Springer; 2023, Vol. 14056:356-366. [doi: 10.1007/978-3-031-48044-7_26]
75. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. Presented at: 36th International Conference on Neural Information Processing Systems; Nov 28 to Dec 9, 2022. [doi: 10.5555/3600270.3602070]
76. Ji Z, Yu T, Xu Y, Lee N, Ishii E, Fung P. Towards mitigating LLM hallucination via self reflection. In: Bouamor H, Pino J, Bali K, editors. Findings of the Association for Computational Linguistics: EMNLP 2023: Association for Computational Linguistics; 2023:1827-1843. [doi: 10.18653/v1/2023.findings-emnlp.123]
77. AuYoung M, Rodriguez Espinosa P, Chen WT, et al. Addressing racial/ethnic inequities in vaccine hesitancy and uptake: lessons learned from the California alliance against COVID-19. J Behav Med 2023 Apr;46(1-2):153-166. [doi: 10.1007/s10865-022-00284-8] [Medline: 35066696]

XSL•FO
RenderX

Abbreviations

LLM: large language model
WHO : World Health Organization

Edited by E Mensah; submitted 10.Feb.2025; peer-reviewed by J Luchman, R Zhang; revised version received 05.Jan.2026; accepted 17.Jan.2026; published 27.Apr.2026.

Please cite as:
Aleksandric A, Dangal A, Nilizadeh S, Mustata Wilson G
Assessment of the Cultural Nuances in COVID-19 Vaccine Uptake Through a Comparative Analysis of English and Spanish Facebook Posts in Tarrant County, Texas: Longitudinal Study
Online J Public Health Inform 2026;18:e72465
URL: https://ojphi.jmir.org/2026/1/e72465
doi:10.2196/72465

© Ana Aleksandric, Anisha Dangal, Shirin Nilizadeh, Gabriela Mustata Wilson. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 27.Apr.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Social Media Use, Professionalism, and Professional Identity Among Medical Students and Early-Career Health Care Graduates in Sudan: Cross-Sectional Survey Study

Fatima Elbasri Abuelgasim Mohammed[1], MBBS; Enas Osama Hassan Omer[2], MBBS; Ammar Elgadi[1], MBBS; Ola Rashid Eltayeb Mamoun[2]; Mohamed Shawgi Yassin Alsayed[3], MBBS; Saad Elsayed[4], MBBS; Ragheda Gamal Abdelrafi Musaad[4], MBBS; Salma Sleak Alam[5], BPharm; Mohamed Adam[6], MBBS; Bayadir Adam Mohamed Adam[6], MBBS; Sara Isam Eldin Ibrahim Elshafie[1], MBBS; Mohanad Mustafa Mohamed Elamin[7], MBBS; Omer Abdallah Mohamedelhassan Mohamed[8], MBBS

[1]Faculty of Medicine, University of Khartoum, Alqast Street, Khartoum, Sudan
[2]Faculty of Medicine, University of Bahri, Khartoum, Sudan
[3]Faculty of Medicine, Al-neelain University, Khartoum, Sudan
[4]Faculty of Medicine, Elrazi University, Khartoum, Sudan
[5]Faculty of Pharmacy, Sudan University of Science and Technology, Khartoum, Sudan
[6]Faculty of Medicine, University of Medical Sciences and Technology, Khartoum, Sudan
[7]National Ribat University, Khartoum, Sudan
[8]Community Medicine Department, Merowe University of Technology, Merowe, Sudan

Corresponding Author:
Fatima Elbasri Abuelgasim Mohammed, MBBS
Faculty of Medicine, University of Khartoum, Alqast Street, Khartoum, Sudan

Abstract

Background: Social media has become deeply embedded in medical training, yet its influence on professionalism and the developing professional identity of medical students and early-career doctors remains insufficiently understood, particularly in non-Western contexts.

Objective: This study aimed to assess social media usage patterns, attitudes toward digital professionalism, related behaviors and experiences, and perceived benefits and risks among medical students and early-career health care graduates in Sudan, and to identify demographic and behavioral factors associated with key online behaviors and outcomes.

Methods: A cross-sectional online survey was conducted from June 2025 to August 2025 among 432 medical students, interns, early graduates, and residents in Sudan, recruited through convenience sampling. A structured, self-administered questionnaire developed specifically for this study assessed social media usage patterns, attitudes toward digital professionalism, related behaviors and experiences, and perceived benefits and risks. Descriptive statistics, chi-square tests, and multivariable logistic regression were performed using R software (version 4.4.2; R Foundation for Statistical Computing).

Results: Of the 432 participants, 226 (52.3%) were female participants, and 343 (79.4%) spent at least 3 hours daily on social media. Attitudes toward digital professionalism were generally positive: 225 (52.1%) participants agreed or strongly agreed that their online presence reflects their professional identity, and 390 (90.3%) considered maintaining a professional image is important. Nearly half of the participants (n=199, 46.1%) had modified or removed posts for professional reasons. The most reported benefits were access to updated medical knowledge (n=321, 74.3%) and career development (n=260, 60.2%), while common concerns involved misinformation (n=287, 66.4%), negative public perception (n=199, 46.1%), and confidentiality breaches (n=198, 45.8%). Female participants were more likely to report heavy daily use of 5 hours or more (109/226, 48.2% vs 70/206, 33.9%; P=.004). In multivariable analysis, male sex (odds ratio 2.12, 95% CI 1.36 - 3.33; P<.001) and posting health care–related content (odds ratio 2.64, 95% CI 1.58 - 4.52; P<.001) independently predicted experiencing a negative outcome on social media, whereas confidence in distinguishing appropriate content showed no significant demographic variation.

Conclusions: Medical students and early-career doctors in Sudan engage heavily with social media and recognize both its value and its risks to professional identity. These findings highlight the need for structured digital professionalism training and culturally sensitive guidelines to support safe and responsible online engagement.

(Online J Public Health Inform 2026;18:e89260) doi:10.2196/89260

XSL•FO
RenderX

KEYWORDS

social media; digital professionalism; professional identity formation; medical students; early-career doctors

Introduction

Background

As digital environments become embedded in the culture of medicine, important questions remain regarding how these online interactions influence the development of a student's professional identity—a hallmark of medical formation. Understanding this dynamic has become a global priority, as educators and regulators attempt to balance the opportunities of online engagement with the preservation of professionalism and public trust.

Professional identity formation (PIF) is the process by which medical students internalize the values, behaviors, and norms of the medical profession and come to "think, act, and feel like a physician" [1]. It is both an individual and a social process, shaped by educational structures, mentorship, institutional culture, and peer interactions. Both PIF frameworks and social identity theory propose that identity develops through an individual's integration into a professional community [1]. Social media has added a new dimension to this process: it reshapes how members of the medical community interact within the framework of professionalism, giving rise to identities that are increasingly shaped by digital visibility and online participation [1,2]. If used effectively, social media platforms such as X (formerly Twitter), LinkedIn, Instagram, and YouTube can facilitate networking and knowledge and experience sharing [2]. In addition, medical professionals can use social media to advocate for patients' rights and raise awareness regarding emerging health issues [2]. Furthermore, social media can be used to create meaningful learning experiences, facilitate collaboration among peers, and foster reflective discussions [3,4]. However, social media platforms can predispose medical students and early-career doctors to ethical dilemmas and reputational damage. Many debates about these platforms' impact on confidentiality, misinformation, and the blurring of boundaries have raised questions about the extent to which individuals should conduct themselves professionally on these platforms [5,6].

Different studies have discussed the connection between social media and professional identity. For example, Garner and O'Sullivan [7] reported that medical students in the United Kingdom expressed concern over practices among their colleagues in using images on Facebook that might impact their professional identity. White et al [8] reported that the majority of students who used Facebook identified content about alcohol or drug use, obesity, criticism of others, and sexual content as unprofessional behaviors, although 44% of them had seen a colleague posting about these negative behaviors. It has become inevitable that medical students and early-career doctors become aware of the implications that social media has on ethics and professionalism [9]. However, a large majority of the existing literature is grounded in Western educational and cultural contexts, wherein norms of self-presentation and professional autonomy differ markedly from those in collectivist societies. In many regions, including the Middle East and North Africa, cultural expectations surrounding modesty, authority, and social reputation interact strongly with perceptions of professionalism. Thus, the negotiation of identity in digital spaces may be considerably more complex, as students must reconcile globalized norms of visibility and advocacy with local ethical frameworks and institutional hierarchies [10]. Taken together, prior studies establish that unprofessional online behavior is common and consequential, and that social media meaningfully shapes professional identity; however, this evidence base offers little evidence from African or Middle Eastern settings, and, to our knowledge, no study to date has examined how medical students and early-career doctors in Sudan perceive the relationship between social media and their professional identity.

Understanding these nuances is important, as PIF is context-dependent. What counts as self-expression or advocacy in one setting may be considered unprofessional or inappropriate in another. In the absence of empirical data that reflect these sociocultural differences, the creation of social media guidelines or educational interventions will fall victim to misalignment with local realities [10].

In addition, it is important to highlight 2 major concerns. First, while many medical institutions globally have issued guidelines on digital professionalism, few provide comprehensive training that helps students and doctors critically reflect on their digital identities [11]. Second, there is limited empirical understanding of how cultural context mediates students' and doctors' online experiences and how these, in turn, influence their developing sense of professionalism.

Digital professionalism needs to be taught not as a set of rules, but as a reflective competency—a practice of judgment, accountability, and self-awareness. Formal curricula, institutional policies, and frameworks for mentoring now need to be developed to help students manage their digital presence responsibly and authentically. Several studies have found that interactive workshops, reflective exercises, and faculty modeling can significantly enhance students' awareness of professional behavior online [11-13].

Objectives

The aim of this study was to investigate the perspectives of medical students and early-career health care graduates in Sudan regarding the influence of social media on medical professional identity. Specifically, the study sought to (1) describe social media usage patterns; (2) assess attitudes toward professionalism and professional identity online; (3) document professionalism-related behaviors, experiences, and perceived benefits and risks; and (4) identify demographic and behavioral factors associated with heavy social media use, health care–related content engagement, confidence in posting appropriateness, and negative online outcomes.

XSL•FO
RenderX

Methods

Study Design and Setting

A cross-sectional, institution-based study was conducted from June 2025 to August 2025 among medical students and early-career health professionals in Sudan.

Study Population and Sampling

We targeted medical students from the first to the final year of study. In addition, we included early-career health professionals (interns, early graduates in practice, and residents) who had graduated within the previous 10 years. Because there is no single, universally accepted definition of the early-career phase, different bodies apply varying cutoffs, we operationalized "early career" as within 10 years of graduation, consistent with Rotenstein et al [14], who characterize the early-career phase as approximately the first decade in which physicians establish themselves in independent practice and develop their professional identity. Convenience sampling was used, and participants were reached through an anonymous online questionnaire (Google Forms). The survey link was open to anyone who received it and was distributed via medical social media groups and WhatsApp groups dedicated mainly to medical students and health care professionals. Participation was voluntary, and no incentives were offered.

Sample Size Calculation

The sample size was calculated using the following formula: $n = (z^2 \times p \times q)/d^2$, where n is the sample size, z is the *z* score (1.96 for a 95% confidence level), p is the expected proportion (0.5), q is 1−p, and d is the desired margin of error (0.05), yielding a minimum required sample size of 384 participants.

Data Collection Tool and Procedure

An anonymous, structured, self-administered questionnaire in English was developed specifically for this study to assess medical students' and early medical graduates' social media use, perceptions of digital professionalism, and experiences related to professional identity. The items were newly constructed by the research team rather than adapted from a previously validated instrument, and their content was informed by the existing literature on digital professionalism and online identity among medical students and health care professionals [2,7,8,12]. The instrument consisted of 6 sections. The first section captured participant demographics (6 items), including age, sex, academic level, and current status.

The *Perceptions of Professionalism and Professional Identity on Social Media* section consisted of 5 attitudinal items. These items assessed participants' views on the professional implications of their social media presence, including whether their online activity reflects their identity as a health care professional, the importance of maintaining a professional image, the perceived influence of social media on public trust, the potential enhancement of professional reputation through online engagement, and their confidence in distinguishing appropriate from inappropriate online content. Because these items represent conceptually related attitudes toward digital professionalism, they were evaluated psychometrically. Responses were coded on ordinal or binary scales according to established scoring rules. The construct validity of the scale was assessed using exploratory factor analysis (EFA; Multimedia Appendix 1), and Cronbach α was subsequently calculated for the initial 5-item scale and for the final 4-item scale retained after factor analysis, with item-total statistics used to evaluate the contribution of each item (Multimedia Appendix 2).

The *Experiences and Behaviors* domain consisted of 7 items that assessed participants' real-world actions and encounters related to social media use within health care contexts. These items included following health care–related pages, posting health care–related content, modifying or removing posts for professional reasons, experiencing negative outcomes related to social media activity, maintaining separate personal and professional accounts, encountering inappropriate content shared by health care professionals, and using social media for various educational or professional purposes. Collectively, these items capture diverse behavioral patterns and experiential events, ranging from usage habits to corrective behaviors and exposure to unprofessional content. Because these items represent independent experiences rather than indicators of a single latent construct, they are not expected to correlate strongly. In addition, the items vary in format, some binary (yes or no), others categorical, and one functioning as a multiresponse checklist, making internal consistency measures such as Cronbach α methodologically inappropriate. Therefore, reliability analysis was not conducted for this domain.

Two additional checklist items were included to assess participants' knowledge and awareness regarding social media use in health care. These questions asked respondents to identify the benefits they associate with using social media as health care students or professionals and the risks they perceive in a health care context. As checklist-style items reflecting knowledge rather than attitudes or behaviors, they were analyzed descriptively and were not included in reliability testing. The survey was reported in accordance with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines (Checklist 1).

The questionnaire was administered via Google Forms across five pages: (1) informed consent; (2) demographic characteristics, including selection of career level (undergraduate or postgraduate); (3) items for medical students (eg, university and academic year); (4) items for practicing doctors (eg, public or private hospital); and (5) the attitude, behavior, and checklist items described above. All items were mandatory, so incomplete questionnaires could not be submitted, and respondents were unable to change their answers after submission. Items and response options were presented in a fixed order for all respondents, without randomization, and no adaptive questioning was used. No technical measures (cookies, IP address checks, or log file analysis) were applied to prevent duplicate entries; instead, data collectors instructed participants to complete the questionnaire only once. No submissions with atypical timestamps were identified. Because the survey link was distributed openly, the number of individuals who received or viewed the invitation is unknown, and viewing and participation rates could not be calculated.

XSL•FO
RenderX

Questionnaire Validity

The questionnaire underwent a multistep validation process to ensure clarity, relevance, and alignment with the study constructs. Face validity was assessed through a pilot test with 14 medical students and early graduates, who evaluated the questionnaire for comprehensibility, wording clarity, and response burden. Minor adjustments were made to item phrasing and formatting based on participant feedback to optimize ease of completion and alignment with respondents' understanding. The 14 pilot participants were excluded from the final analytic sample (N=432) to avoid potential bias.

Ethical Considerations

Electronic informed consent was collected from all study participants before they could access the questionnaire. Ethical approval was obtained from the research ethics committee of Nafeer Elafia Hospital, the Northern State Ministry of Health, Northern State, Sudan (reference number 300101; approved on January 30, 2025). The study was conducted according to the principles of the Declaration of Helsinki, and all responses were collected anonymously.

Statistical Analysis

Data were analyzed using R software (version 4.4.2; R Foundation for Statistical Computing). Descriptive statistics were used to summarize demographic characteristics, social media usage patterns, and responses to questionnaire items, with results reported as frequencies and percentages.

For the attitudinal section assessing perceptions of professionalism and professional identity on social media, responses were numerically coded according to the predefined scoring scheme, and a total professionalism attitude score was generated. Items in the *Experiences and Behaviors* domain and the *Benefits and Risks Awareness* domain were treated as independent categorical indicators and summarized descriptively, as they did not represent a single latent construct suitable for scale formation.

The construct validity of the professionalism attitude scale was assessed using EFA. Data suitability was confirmed with the Kaiser-Meyer-Olkin measure of sampling adequacy and the Bartlett test of sphericity; the number of factors to retain was determined by parallel analysis, and factors were extracted using the minimum residual (MINRES) method with oblimin rotation. Full details are provided in Multimedia Appendix 1.

Bivariate analyses using Pearson chi-square tests were conducted to explore associations between 3 primary demographic characteristics (sex, age group, and current status) and key social media behaviors and perceptions: heavy usage (≥5 h/d), following health care pages, posting health care content, confidence in posting appropriateness, and experiencing negative outcomes. Chi-square statistics are reported with their degrees of freedom. Because these bivariate tests were exploratory, unadjusted *P* values are reported and no formal correction for multiple testing was applied; instead, multivariable logistic regression was used to control for confounding and to reduce reliance on unadjusted bivariate comparisons.

Multivariable logistic regression models were then fitted for each of the same 5 binary outcomes to identify independent demographic and behavioral predictors. All models included sex, age (dichotomized as ≤25 y vs >25 y), current status (with medical student as the reference), and maintenance of separate personal and professional accounts as predictors; relevant behavioral covariates (daily usage, following health care pages, confidence in post appropriateness, and posting health care content) were additionally included when conceptually appropriate for the outcome being modeled. Results are reported as odds ratios (ORs) with 95% CIs. A significance level of *P* less than .05 was used for all statistical tests.

Results

Demographic Characteristics

A total of 432 participants were included in the study. Most were aged 26 to 30 years (n=197, 45.6%), followed by 21 to 25 years (n=181, 41.9%), with smaller proportions in the younger and older age groups. Females represented 52.3% (n=226) of the sample. Participants' professional status included medical students (n=135, 31.3%), interns (n=110, 25.5%), early graduates in practice (n=95, 21.9%), and residents (n=92, 21.3%). Among the 135 medical students, the largest representation came from Bahri University (n=45, 33.3%), followed by Elrazi University (n=17, 12.6%), Khartoum University (n=13, 9.6%), and National University (n=11, 8.1%). Students were distributed across academic years, with the highest proportions in the fifth (n=40, 29.6%) and sixth (n=41, 30.4%) years (Table 1).

XSL•FO
RenderX

Table . Demographic characteristics of study participants (N=432).

Characteristic	Values, n (%)
Sex	
Female	226 (52.3)
Male	206 (47.7)
Age group (y)	
16 - 20	14 (3.2)
21 - 25	181 (41.9)
26 - 30	197 (45.6)
31 - 35	33 (7.6)
36 - 40	7 (1.6)
Current status	
Medical student	135 (31.3)
Intern	110 (25.5)
Early graduate in practice	95 (21.9)
Resident	92 (21.3)
University (medical students, n=135)	
Ahfad University	5 (3.7)
Al-Neelain University	2 (1.5)
Alzaim Alazhary University	1 (0.7)
Bahri University	45 (33.3)
Dongola University	2 (1.5)
El-Fagr University	2 (1.5)
Elrazi University	17 (12.6)
Gezira University	1 (0.7)
Ibn Sina University	1 (0.7)
Kassala University	1 (0.7)
Khartoum University	13 (9.6)
Kordofan University	4 (2.9)
National Ribat University	8 (5.9)
National University	11 (8.1)
Sudan International University	6 (4.4)
Sudan University of Science and Technology	6 (4.4)
University of El Imam El Mahdi	2 (1.5)
University of Medical Science and Technology	8 (5.9)
Academic year (medical students, n=135)	
First	5 (3.7)
Second	4 (2.9)
Third	16 (11.9)
Fourth	29 (21.5)
Fifth	40 (29.6)
Sixth	41 (30.4)

XSL•FO
RenderX

Social Media Usage Patterns

Of the 432 participants, WhatsApp, Instagram (Meta Platforms, Inc.), and Facebook (Meta Platforms, Inc.) were the most commonly used platforms, reported by 322 (74.5%), 276 (63.9%), and 211 (48.8%) participants, respectively. TikTok (ByteDance Ltd) and X (formerly Twitter) were less frequently used (Figure S1 in Multimedia Appendix 3). Daily usage was high, with 179 (41.4%) participants spending at least 5 hours per day on social media and another 164 (37.9%) spending 3 to 4 hours per day. Most participants followed health care–related pages (n=381, 88.2%) and reported varying frequencies of posting health care–related content: rarely (n=139, 32.2%), sometimes (n=127, 29.4%), never (n=119, 27.5%), and frequently (n=47, 10.9%; Table 2). Participants primarily used social media for educational purposes (n=335, 77.5%), as well as for entertainment (n=287, 66.4%) and networking (n=241, 55.8%). About one-quarter (n=115, 26.6%) used social media for patient communication or public health advocacy (Figure S2 in Multimedia Appendix 3).

Table . Social media usage patterns among participants (N=432).

Items and responses	Values, n (%)
Average daily time spent on social media (h)	
Less than 1	14 (3.2)
1 - 2	75 (17.4)
3 - 4	164 (37.9)
5 or more	179 (41.4)
Follows health care–related pages, professionals, or organizations	
Yes	381 (88.2)
No	51 (11.8)
Posts health care–related content (eg, medical news, infographics, or opinions)	
Frequently	47 (10.9)
Sometimes	127 (29.4)
Rarely	139 (32.2)
Never	119 (27.5)

Reliability and Construct Validity of the Professionalism Attitude Scale

The EFA supported a unidimensional structure: the single extracted factor explained 26.0% of the total variance with excellent model fit (χ^2=2.2, *P*=.82; Tucker-Lewis Index=1.02; root mean square error of approximation=0.0, 90% CI 0.00 - 0.04). Four of the 5 items showed salient loadings (≥0.40) on this factor, whereas the behavioral item on modifying or removing a post did not load on the attitudinal construct and was excluded from the final validated scale (Multimedia Appendix 1). Following the EFA, internal consistency was assessed. The initial 5-item scale demonstrated a raw Cronbach α of 0.60, and item-total statistics supported the EFA findings, indicating that removal of the behavioral item would improve the standardized internal consistency. The final validated 4-item scale demonstrated moderate reliability (raw Cronbach α=0.60 and standardized Cronbach α=0.62), which is considered acceptable for a brief scale capturing the underlying professionalism attitude construct (Multimedia Appendix 2).

Attitudes Toward Professionalism and Professional Identity

Of the 432 participants, attitudes toward digital professionalism were generally positive. More than half agreed that their social media presence reflects their identity as current or future health care professionals, including 87 (20.1%) who strongly agreed and 138 (31.9%) who agreed, while only 74 (17.1%) disagreed or strongly disagreed. Most participants recognized the importance of maintaining a professional online image, with 199 (46.1%) rating it as very important and 191 (44.2%) as somewhat important. Nearly half of the participants (n=199, 46.1%) had modified or removed a post to align with professional expectations. Perceptions of social media's broader impact were also favorable: 302 (69.9%) participants agreed or strongly agreed that social media influences public trust in health care professionals, and an equal proportion (n=302, 69.9%) believed it enhances their professional reputation (Table 3).

XSL•FO
RenderX

Table . Attitudes toward professional identity and professionalism on social media (N=432).

Items and responses	Values, n (%)
Social media presence reflects professional identity	
Strongly agree	87 (20.1)
Agree	138 (31.9)
Neutral	133 (30.8)
Disagree	50 (11.6)
Strongly disagree	24 (5.6)
Importance of professional image online	
Very important	199 (46.1)
Somewhat important	191 (44.2)
Not important	42 (9.7)
Social media influences public trust	
Strongly agree	110 (25.5)
Agree	192 (44.4)
Neutral	108 (25)
Disagree	16 (3.7)
Strongly disagree	6 (1.4)
Engaging on social media enhances professional reputation	
Yes	302 (69.9)
No	130 (30.1)
Modified or removed a post for professional reasons	
Yes	199 (46.1)
No	233 (53.9)

Experiences and Online Behaviors

Among the 432 participants, experiences related to professionalism on social media varied. Over one-third (n=153, 35.4%) had experienced a negative outcome related to their social media activity. Confidence in distinguishing appropriate from inappropriate posts was generally high, with 261 (60.4%) participants expressing confidence, though 131 (30.3%) reported only sometimes feeling confident and 40 (9.3%) reported no confidence. Exposure to unprofessional content was relatively common: 142 (32.9%) participants had encountered a health care professional posting inappropriate material online. Regarding account management practices, 128 (29.6%) participants maintained separate personal and professional accounts, 175 (40.5%) did not, and 129 (29.9%) were planning to do so (Table 4).

XSL•FO
RenderX

Table . Experiences and online behaviors related to social media use (N=432).

Items and responses	Values, n (%)
Experienced a negative outcome due to social media	
Yes	153 (35.4)
No	279 (64.6)
Confidence in distinguishing appropriate posts	
Yes	261 (60.4)
Sometimes	131 (30.3)
No	40 (9.3)
Encountered a health care professional posting inappropriate content	
Yes	142 (32.9)
No	290 (67.1)
Maintains separate personal and professional accounts	
Yes	128 (29.6)
No	175 (40.5)
Planning to	129 (29.9)

Perceived Benefits and Risks of Social Media Use in a Health Care Context

Among the 432 participants, several professional advantages of using social media were identified. The most frequently reported benefit was access to updated medical knowledge (n=321, 74.3%), followed by opportunities for career development (n=260, 60.2%), and professional networking (n=235, 54.4%). A substantial proportion also highlighted the role of social media in public health advocacy (n=177, 40.9%) and enhancing visibility or recognition as a health care professional (n=150, 34.7%). These findings indicate that participants largely view social media as a platform that supports both learning and career progression (Figure S3 in Multimedia Appendix 3).

Among the 432 participants, a range of concerns regarding the use of social media in health care was reported. The most frequently cited risk was the spread of misinformation (n=287, 66.4%), followed by fears of negative public perception (n=199, 46.1%), and potential breaches of patient confidentiality (n=198, 45.8%). Additional concerns included blurred personal-professional boundaries (n=142, 32.9%), job-related consequences (n=111, 25.7%), and experiences with or fears of cyberbullying or trolling (n=101, 23.4%). These results illustrate a balanced understanding among participants of both the professional opportunities and the potential hazards associated with social media use (Figure S4 in Multimedia Appendix 3).

Demographic Associations With Social Media Behaviors and Outcomes

Bivariate analyses using Pearson chi-square tests examined associations between respondent characteristics (sex, age, and current professional status) and social media behaviors, confidence, and experiences (Table 5).

XSL•FO
RenderX

Table . Associations between demographic characteristics and key social media behaviors (N=432).

Characteristic	≥5 hours/day	Follows health care pages	Posts health care content	Confident in appropriateness	Negative outcome
Sex					
Female (n=226), n (%)	109 (48.2)	201 (88.9)	170 (75.2)	127 (56.2)	68 (30.1)
Male (n=206), n (%)	70 (33.9)	180 (87.4)	143 (69.4)	134 (65.0)	85 (41.3)
Chi-square (*df*); *P* value[a]	8.44 (1); .004	0.12 (1); .73	1.54 (1); .22	3.17 (1); .08	5.40 (1); .02
Age group (y)					
16 - 20 (n=14), n (%)	4 (28.6)	10 (71.4)	8 (57.1)	9 (64.3)	4 (28.6)
21 - 25 (n=181), n (%)	72 (39.8)	160 (88.4)	118 (65.2)	110 (60.8)	67 (37.0)
26 - 30 (n=197), n (%)	92 (46.7)	176 (89.3)	157 (76.7)	114 (57.9)	67 (34.0)
31 - 35 (n=33), n (%)	7 (21.2)	28 (84.8)	26 (78.8)	23 (69.7)	14 (42.4)
36 - 40 (n=7), n (%)	4 (57.1)	7 (100)	4 (57.1)	5 (71.4)	1 (14.3)
Chi-square (*df*); *P* value[a]	9.68 (4); .046	5.33 (4); .26	13.09 (4); .01	2.18 (4); .70	2.73 (4); .60
Current status					
Medical student (n=135), n (%)	55 (40.7)	120 (88.9)	89 (65.9)	73 (54.1)	49 (36.3)
Intern (n=110), n (%)	49 (44.5)	98 (89.1)	83 (75.5)	74 (67.3)	35 (31.8)
Early graduate (n=95), n (%)	40 (42.1)	84 (88.4)	71 (74.7)	58 (61.1)	34 (35.8)
Resident (n=92), n (%)	35 (38.0)	79 (85.9)	70 (76.1)	56 (60.9)	35 (38.0)
Chi-square (*df*); *P* value[a]	0.92 (3); .82	0.63 (3); .89	4.24 (3); .24	4.46 (3); .22	0.95 (3); .81

[a]Pearson chi-square test; values shown are the chi-square statistic (df) followed by the *P* value.

Daily social media use varied significantly across demographic factors. A significant association was observed between sex and the duration of daily social media use (χ^2_1 =8.44; *P*=.004): a higher proportion of female respondents (109/226, 48.2%) reported using social media for 5 or more hours per day compared with male respondents (70/206, 33.9%). Daily usage was also significantly associated with age (χ^2_4 =9.68; *P*=.046), with the highest proportion of heavy users (≥5 h/d) concentrated among respondents aged 26 to 30 years (92/197, 46.7%). Current professional status was not significantly associated with daily usage duration (*P*=.82).

The vast majority of the sample engaged with professional content, with 381 of 432 participants (88.2%) reporting that they follow health care pages. This behavior was consistent across the cohort, showing no significant differences based on sex, age, or current status (*P*>.05 for all). In contrast, active content creation was influenced by age. While there were no significant differences in posting behavior by sex (*P*=.22) or current status (*P*=.24), posting health care–related content was significantly associated with age (χ^2_4 =13.09; *P*=.01): respondents aged 26 to 30 years were the most active creators, with 157 of 197 (79.7%) reporting that they post health care content, compared with 118 of 181 (65.2%) of those aged 21 to 25 years.

Respondents' confidence in the appropriateness of their posts did not show any statistically significant associations across demographic groups, although male respondents reported slightly higher confidence (134/206, 65.0%) compared with female respondents (127/226, 56.2%; χ^2_1 =3.17; *P*=.08). However, when examining adverse digital experiences, a significant association with sex was identified (χ^2_1 =5.40; *P*=.02): male respondents were more likely to report having experienced a negative outcome on social media (85/206, 41.3%) compared with their female counterparts (68/226, 30.1%). Neither age (*P*=.60) nor current professional status (*P*=.81) was significantly associated with experiencing negative digital outcomes.

XSL•FO
RenderX

Multivariable Predictors of Social Media Behaviors and Outcomes

Multivariable logistic regression models examined the demographic and behavioral predictors of 5 distinct social media outcomes: heavy usage (≥5 h/d), following health care pages, posting health care content, confidence in the appropriateness of posting, and experiencing negative outcomes (Table 6).

Table . Multivariable logistic regression analysis of factors associated with social media behaviors and outcomes (N=432)[a].

Predictor	≥5 hours/day, OR[b] (95% CI); *P* value	Follows pages, OR (95% CI); *P* value	Posts content, OR (95% CI); *P* value	Confident, OR (95% CI); *P* value	Negative outcome, OR (95% CI); *P* value
Sex (reference: female)					
Male	0.54 (0.36 - 0.81); .003	0.86 (0.47 - 1.58); .63	0.66 (0.41 - 1.04); .08	1.38 (0.93 - 2.07); .11	2.12 (1.36 - 3.33); <.001
Age (reference: ≤25 y)					
>25 years	1.31 (0.82 - 2.08); .26	1.46 (0.73 - 2.91); .28	2.03 (1.19 - 3.49); .01	0.73 (0.46 - 1.17); .20	0.76 (0.46 - 1.25); .28
Current status (reference: medical student)					
Early graduate	1.04 (0.58 - 1.88); .89	0.84 (0.34 - 2.10); .70	1.21 (0.62 - 2.39); .58	1.46 (0.81 - 2.64); .21	0.92 (0.48 - 1.72); .78
Intern	1.18 (0.67 - 2.08); .58	0.91 (0.38 - 2.21); .83	1.24 (0.64 - 2.40); .53	1.98 (1.11 - 3.59); .02	0.82 (0.44 - 1.53); .54
Resident	0.88 (0.46 - 1.68); .71	0.63 (0.24 - 1.64); .34	1.25 (0.59 - 2.71); .56	1.47 (0.77 - 2.80); .24	1.04 (0.52 - 2.07); .91
Separate accounts (reference: no)					
Planning to	1.12 (0.70 - 1.80); .63	1.26 (0.61 - 2.69); .54	1.25 (0.73 - 2.18); .42	1.03 (0.65 - 1.65); .90	2.03 (1.20 - 3.47); .008
Yes	1.02 (0.63 - 1.64); .94	1.00 (0.50 - 2.03); *P*≥.99	1.52 (0.87 - 2.69); .15	1.75 (1.08 - 2.86); .02	4.13 (2.47 - 7.00); <.001
Uses social media ≥5 hours/day (reference: no)					
Yes	—[c]	0.73 (0.40 - 1.33); .30	—	—	1.29 (0.83 - 2.00); .25
Follows health care pages (reference: no)					
Yes	—	—	5.71 (3.03 - 11.10); <.001	—	—
Confident in posts (reference: no)					
Yes	—	—	1.76 (1.10 - 2.80); .02	—	0.74 (0.47 - 1.15); .18
Posts health care content (reference: no)					
Yes	—	—	—	—	2.64 (1.58 - 4.52); <.001

[a]Each column represents a separate multivariable model adjusted for all predictors, with values shown in that column.

[b]OR: odds ratio.

[c]Variable was not included as a predictor in that specific regression model.

When examining daily social media time, male participants were significantly less likely than female participants to report use of 5 hours per day or more (OR 0.54, 95% CI 0.36 - 0.81; *P*=.003). Age, current training status, and the separation of personal and professional accounts were not significant predictors of heavy usage. The model predicting whether participants followed health care–related pages yielded no statistically significant demographic or behavioral predictors.

Several factors significantly predicted the likelihood of posting health care–related content. Participants older than 25 years had twice the odds of posting content compared with those aged 25 years and younger (OR 2.03, 95% CI 1.19 - 3.49; *P*=.01). Behavioral factors were also strong predictors: those who followed health care pages were nearly 6 times more likely to post their own content (OR 5.71, 95% CI 3.03 - 11.10; *P*<.001), and individuals who reported confidence in the appropriateness of their posts had significantly higher odds of actively posting content (OR 1.76, 95% CI 1.10 - 2.80; *P*=.02).

Current training status and account management strategies were significantly associated with confidence in the appropriateness of social media posts. Compared with medical students, interns were nearly twice as likely to report feeling confident (OR 1.98, 95% CI 1.11 - 3.59; *P*=.02). Participants who actively maintained separate personal and professional accounts also demonstrated significantly higher odds of confidence than those who did not separate their accounts (OR 1.75, 95% CI 1.08 - 2.86; *P*=.02).

The model identified multiple significant predictors of experiencing negative outcomes related to social media use. Male participants had more than twice the odds of reporting a negative outcome compared with their female counterparts (OR 2.12, 95% CI 1.36 - 3.33; *P*<.001). Account separation also

XSL•FO
RenderX

played a substantial role: paradoxically, participants who already maintained separate accounts had more than 4 times the odds of experiencing a negative outcome (OR 4.13, 95% CI 2.47 - 7.00; *P*<.001), and those planning to separate their accounts had twice the odds (OR 2.03, 95% CI 1.20 - 3.47; *P*=.008) compared with those who did not separate them. Finally, posting health care content was a strong independent predictor, significantly increasing the likelihood of experiencing a negative outcome (OR 2.64, 95% CI 1.58 - 4.52; *P*<.001).

Discussion

Principal Findings

In this cross-sectional survey of 432 medical students and early-career doctors in Sudan, we found high levels of social media use, with 179 (41.4%) participants online for at least 5 hours daily, alongside generally positive attitudes toward digital professionalism, growing awareness of the professional implications of online activity, and substantial gaps in training and institutional guidance. Female participants reported heavier daily use, whereas male participants and those posting health care–related content had significantly higher odds of experiencing negative online outcomes. To our knowledge, this is the first study to examine digital professionalism and professional identity among medical students and early-career doctors in Sudan. It advances the existing literature, which has been largely descriptive and grounded in Western educational contexts [7,8,10], in 2 ways: by combining a psychometrically evaluated attitude scale with multivariable modeling of the predictors of key online behaviors and outcomes and by providing empirical evidence from a collectivist, resource-constrained setting where cultural norms around reputation and hierarchy shape how professional identity is negotiated online.

Comparison With Prior Work

Our finding of social media platforms usage for at least 5 hours per day aligns with the findings of Faihs et al [15], who reported an increasing interest among medical students in learning more about digital technologies, reflecting their immersion in digital ecosystems. Moreover, Gagnon and Sabus [16] reflected on the increase in the use of social media by health care providers and the need to propose guidelines for social media usage by physical therapy professionals, which indicates the need to maintain patients' privacy and adhere to the principles of medical ethics even when social media is used for educational purposes. The study participants, including both undergraduates and postgraduates, reported growing trends toward the use of social media in informal learning activities besides knowledge sharing and networking. In contrast, Sterling et al [4] reported that there is inconclusive evidence on the impact of social media on graduate medical education and professionalism despite its common use. Around half of the study participants (225/432, 52.1%) agreed that their social media accounts reflect their professional identity, emphasizing that digital spaces have become essential in shaping identity and supporting the findings of Wu et al [17] that social media shapes professional identity among medical students through informational and emotional support. Out of 432 participants, 199 (46.1%) had incidents where they removed or edited their social media posts because of considerations of their professional role, while 233 (53.9%) had not. These differences indicate differences in the perception and management of the professional implications of their online engagement. While the reasons for these behaviors were not directly measured in this study, one possible explanation is that those who removed or edited their content had greater awareness of the professional and reputational consequences of their posts, while those who did not may have perceived their content as appropriate or were less aware of potential professional consequences. While this interpretation remains speculative, the variation in practices is consistent with an absence of systematic training or institutional policies on professional conduct on social media. Ezeilo et al [18] reported that social media has been widely used in health advocacy and awareness-raising, while only 115 (26.6%) participants used social media for advocacy and public health promotion. This suggests that medical students and early-career doctors are not fully engaging with social media for advocacy, perhaps due to perceived gaps in social accountability, training, or confidence. Interestingly, 302 (69.9%) participants believed that social media engagement can improve professional reputation, consistent with the study by Ferrell and Campos-Castillo [19], which suggested that tackling misinformation on X and engaging with the public can improve credibility. Participants also expressed concerns regarding social media practices that negatively impact them as medical students and early-career doctors, such as the spread of misinformation (n=287, 66.4%), negative public perception (n=199, 46.1%), and breaches of patient confidentiality (n=198, 45.8%). These concerns reflect longstanding issues outlined in the literature warning that digital platforms can amplify unprofessional conduct, expose trainees to reputational harm, and diminish public trust in the medical profession [20]. Although many participants reported confidence in distinguishing professional from unprofessional content, the confidence level was not associated with the level of education or training. Because curricular exposure to digital professionalism was not directly assessed in this study, this interpretation is hypothetical; nevertheless, the absence of a gradient across training stages would be consistent with limited formal education on the influence of social media on medical professionalism. Approximately one-third of respondents (n=142, 32.9%) reported encountering inappropriate content shared by health care professionals, a finding consistent with Kitsis et al [12], who reported that medical students engaged in sharing unprofessional content on social media.

Implications for Medical Education and Practice

The findings carry important implications for medical education and professional regulation. There is a need for structured curricula that address topics such as digital professionalism, ethical online behavior, management of digital identity, and responsible engagement in public communication. In addition, given the frequent use of social media for medical learning by participants, educators and institutions could leverage social media platforms to disseminate evidence-based content and create supportive online learning communities. Clear institutional guidelines, mentorship, and support systems to address cyberbullying and negative online experiences would

XSL•FO
RenderX

also help in promoting safe and professional use of digital environments.

Limitations

There are several limitations to this study that should be considered. The cross-sectional design limits the ability to follow changes in attitudes and behaviors over time. Self-reported data may be subject to recall bias and the desire to provide socially acceptable answers, particularly around professionalism. The sample, while large, was a convenience sample obtained through an online questionnaire; this form of sampling is prone to self-selection bias and is likely to overrepresent those who are already active social media users, which may inflate estimates of usage intensity and engagement and limit the generalizability of the results to all medical students and early-career health professionals in Sudan. In addition, the inclusion of professionals who graduated up to 10 years previously introduces some heterogeneity in experience, although current status (student, intern, early graduate, or resident) was accounted for in the stratified and multivariable analyses. Despite these limitations, the study offers valuable insights into the perceptions of emerging health care professionals regarding the junction of social media and professional identity. Future research should investigate longitudinal changes in digital behavior and professionalism throughout medical training.

Conclusions

This study highlights the growing significance of social media in the daily lives and professional journeys of medical students and early-career doctors in Sudan. Participants acknowledged the educational and networking opportunities offered by these platforms, while also being aware of the ethical risks and possible impacts on their professional identities. Although most respondents demonstrated awareness of professional norms by editing or deleting inappropriate posts and recognizing the effect of online behavior on public confidence, practices varied, indicating significant gaps in training and institutional support. The noted sex differences, with female participants reporting higher daily usage and male participants experiencing more frequent negative online experiences, highlight the complex navigation of professional identity within digital spaces. The results demonstrate the value of systematic and context-sensitive education in digital professionalism within medical curricula. This kind of education needs to be more comprehensive than rule-based education and should encompass reflective judgment, responsible online behavior, and confidence in the management of one's digital identity. Therefore, institutions can prepare future health care professionals to participate in digital environments safely, ethically, and meaningfully.

Acknowledgments

Claude (latest version; Anthropic) and ChatGPT 5.5 (OpenAI) were used to assist with manuscript editing and revision to improve phrasing, spelling, the literature review, and structural organization according to journal guidelines. All substantive content, original ideas, data, analyses, and references were generated by the authors. AI tools did not generate research findings, conduct analyses, or contribute to the intellectual content of the work. All references cited were verified by the authors for accuracy and relevance.

Funding

All authors declared that they had insufficient funding to support open-access publication of this manuscript, including from their affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee for the publication of this article.

Data Availability

The data are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: EOHO, FEAM (coleadership and leadership)
Data curation: SSA, MA, BAMA, SIEIE, MMME, OREM, MSYA, SE, RGAM (data collection)
Formal analysis: AE
Investigation: FEAM
Writing – original draft: EOHO, AE, SSA, MA, BAMA, SIEIE, MMME, OREM, MSYA, SE, RGAM, OAMM, FEAM
Writing – review & editing: OAMM, FEAM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exploratory factor analysis and construct validity of the professionalism attitude scale.
[DOCX File, 15 KB - ojphi_v18i1e89260_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Internal consistency reliability analysis of the professionalism attitude scale.
[DOCX File, 16 KB - ojphi_v18i1e89260_app2.docx]

Multimedia Appendix 3
Supplementary figures showing participants' perceived risks and benefits of social media use in health care, purposes of social media use, and regularly used social media platforms (N=432).
[DOCX File, 397 KB - ojphi_v18i1e89260_app3.docx]

Checklist 1
CHERRIES checklist.
[DOCX File, 18 KB - ojphi_v18i1e89260_app4.docx]

References

1. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. Reframing medical education to support professional identity formation. Acad Med 2014 Nov;89(11):1446-1451. [doi: 10.1097/ACM.0000000000000427] [Medline: 25054423]
2. Chretien KC, Greysen SR, Chretien JP, Kind T. Online posting of unprofessional content by medical students. JAMA 2009 Sep 23;302(12):1309-1315. [doi: 10.1001/jama.2009.1387] [Medline: 19773566]
3. Kind T, Patel PD, Lie D, Chretien KC. Twelve tips for using social media as a medical educator. Med Teach 2014 Apr;36(4):284-290. [doi: 10.3109/0142159X.2013.852167] [Medline: 24261897]
4. Sterling M, Leung P, Wright D, Bishop TF. The use of social media in graduate medical education: a systematic review. Acad Med 2017 Jul;92(7):1043-1056. [doi: 10.1097/ACM.0000000000001617] [Medline: 28225466]
5. Farnan JM, Snyder Sulmasy L, Worster BK, et al. Online medical professionalism: patient and public relationships: policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627. [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]
6. Cain J, Scott DR, Akers P. Pharmacy students' Facebook activity and opinions regarding accountability and e-professionalism. Am J Pharm Educ 2009 Oct 1;73(6):104. [doi: 10.5688/aj7306104] [Medline: 19885073]
7. Garner J, O'Sullivan H. Facebook and the professional behaviours of undergraduate medical students. Clin Teach 2010 Jun;7(2):112-115. [doi: 10.1111/j.1743-498X.2010.00356.x] [Medline: 21134159]
8. White J, Kirwan P, Lai K, Walton J, Ross S. Have you seen what is on Facebook?'The use of social networking software by healthcare professions students. BMJ Open 2013;3(7):e003013. [doi: 10.1136/bmjopen-2013-003013] [Medline: 23883886]
9. Chretien KC, Kind T. Social media and clinical care: ethical, professional, and social implications. Circulation 2013 Apr 2;127(13):1413-1421. [doi: 10.1161/CIRCULATIONAHA.112.128017] [Medline: 23547180]
10. Al-Rumayyan A, Van Mook W, Magzoub ME, et al. Medical professionalism frameworks across non-Western cultures: a narrative overview. Med Teach 2017 Apr;39(sup1):S8-S14. [doi: 10.1080/0142159X.2016.1254740] [Medline: 28417688]
11. Protogiros D, Tsitsi T, Cloconi C, et al. Mapping and assessing existing digital skills training programs for health care professionals and health managers in Europe: expert-based survey for investigating the landscape, accessibility, and effectiveness of training initiatives. J Med Internet Res 2025 Aug 29;27:e71657. [doi: 10.2196/71657] [Medline: 40882225]
12. Kitsis EA, Milan FB, Cohen HW, et al. Who's misbehaving? Perceptions of unprofessional social media use by medical students and faculty. BMC Med Educ 2016 Feb 18;16:67. [doi: 10.1186/s12909-016-0572-x] [Medline: 26887561]
13. Mostaghimi A, Crotty BH. Professionalism in the digital age. Ann Intern Med 2011 Apr 19;154(8):560-562. [doi: 10.7326/0003-4819-154-8-201104190-00008] [Medline: 21502653]
14. Rotenstein L, Harry E, Shanafelt T. The early career phase: fostering physician well-being over the career life cycle. Mayo Clin Proc 2025 Oct;100(10):1836-1845. [doi: 10.1016/j.mayocp.2025.05.025] [Medline: 40900071]
15. Faihs V, Figalist C, Bossert E, Weimann K, Berberat PO, Wijnen-Meijer M. Medical students and their perceptions of digital medicine: a question of gender? Med Sci Educ 2022 Oct;32(5):941-946. [doi: 10.1007/s40670-022-01594-x] [Medline: 36276758]
16. Gagnon K, Sabus C. Professionalism in a digital age: opportunities and considerations for using social media in health care. Phys Ther 2015 Mar;95(3):406-414. [doi: 10.2522/ptj.20130227] [Medline: 24903111]
17. Wu M, Yan J, Qiao C, Yan C. Impact of concurrent media exposure on professional identity: cross-sectional study of 1087 medical students during long COVID. J Med Internet Res 2024 Oct 17;26:e50057. [doi: 10.2196/50057] [Medline: 39418080]
18. Ezeilo CO, Leon N, Jajodia A, Han HR. Use of social media for health advocacy for digital communities: descriptive study. JMIR Form Res 2023 Nov 14;7:e51752. [doi: 10.2196/51752] [Medline: 37962914]
19. Ferrell D, Campos-Castillo C. Factors affecting physicians' credibility on Twitter when sharing health information: online experimental study. JMIR Infodemiology 2022;2(1):e34525. [doi: 10.2196/34525] [Medline: 37113807]

XSL•FO
RenderX

20. Guraya SS, Yusoff MSB, Rashid-Doubell F, et al. Changing professional behaviors in the digital world using the medical education e-professionalism (MEeP) framework—a mixed methods multicentre study. Front Med (Lausanne) 2022;9:846971. [doi: 10.3389/fmed.2022.846971] [Medline: 35425778]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
EFA: exploratory factor analysis
MINRES: minimum residual
OR: odds ratio
PIF: professional identity formation

Edited by E Mensah; submitted 27.Dec.2025; peer-reviewed by A Bethanabatla, C Agbasiere; revised version received 08.Jul.2026; accepted 11.Jul.2026; published 25.Aug.2026.

Please cite as:
Mohammed FEA, Omer EOH, Elgadi A, Mamoun ORE, Alsayed MSY, Elsayed S, Musaad RGA, Alam SS, Adam M, Adam BAM, Elshafie SIEI, Elamin MMM, Mohamed OAM
Social Media Use, Professionalism, and Professional Identity Among Medical Students and Early-Career Health Care Graduates in Sudan: Cross-Sectional Survey Study
Online J Public Health Inform 2026;18:e89260
URL: https://ojphi.jmir.org/2026/1/e89260
doi:10.2196/89260

© Fatima Elbasri Abuelgasim Mohammed, Enas Osama Hassan Omer, Ammar Elgadi, Ola Rashid Eltayeb Mamoun, Mohamed Shawgi Yassin Alsayed, Saad Elsayed, Ragheda Gamal Abdelrafi Musaad, Salma Sleak Alam, Mohamed Adam, Bayadir Adam Mohamed Adam, Sara Isam Eldin Ibrahim Elshafie, Mohanad Mustafa Mohamed Elamin, Omer Abdallah Mohamedelhassan Mohamed. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 25.Aug.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Topic and Sentiment Trends in Semaglutide Discussions on X: Subpopulation-Based Longitudinal Analysis

Parisa Momeni[1*], MSCS; Gabriel Laverghetta[1*], MSCS; Jay Ligatti[1], PhD; Lingyao Li[2], PhD

[1]Bellini College of Artificial Intelligence, Cybersecurity and Computing, University of South Florida, Tampa, FL, United States

[2]School of Information, University of South Florida, Tampa, FL, United States

[*]these authors contributed equally

Corresponding Author:
Parisa Momeni, MSCS
Bellini College of Artificial Intelligence, Cybersecurity and Computing
University of South Florida
4202 E. Fowler Avenue
Tampa, FL, 33620
United States
Phone: 1 8137090539
Email: parisamomeni@usf.edu

Abstract

Background: User experience has a significant impact on pharmaceutical drug effectiveness. Social media platforms like X (formerly Twitter) have become prominent spaces where individuals share their medication-related experiences, especially with widely marketed drugs such as semaglutide. Despite the large volume of conversation, a comprehensive understanding of how various user subpopulations engage with semaglutide-related discussions remains underdeveloped.

Objective: This study aims to explore how semaglutide is perceived and discussed across different X user groups. Within these user groups, we investigate (1) the evolution of sentiment patterns toward semaglutide and (2) the evolution and prevalence of semaglutide-related discussion topics.

Methods: We prepared a dataset consisting of 859,751 X posts (tweets) pertaining to semaglutide, along with related metadata, that were posted between July 2021 and April 2024. We apply sentiment analysis and topic modeling to the collected posts and analyze the sentiment patterns and topics within specific user subpopulations and time periods.

Results: Our analysis reveals a mean sentiment score of –0.24 (SD 0.669) across all posts, with all user subpopulations experiencing a decline in sentiment during the study period. User discussions focus on semaglutide's applications in weight loss and potential side effects, along with economic factors and celebrity/political influence. We also uncover differences in sentiment and discussion topics across user subpopulations. Notably, organizational accounts consistently express less negative sentiment (mean −0.04, SD 0.542) than individuals (mean −0.28, SD 0.605), with a statistically significant difference ($P<.001$), particularly in discussions related to drug efficacy and regulatory concerns. Interrupted time-series analysis shows a marked decrease in sentiment during the November 2022-January 2023 period, coinciding with regulatory announcements about potential adverse effects. In addition, we observe gender-based variations, such as a greater prevalence of discussions involving celebrities and politicians within female user posts (8368/39,786, 21%) compared to male user posts (8087/46,133, 17.5%), and male users expressing more positive sentiment.

Conclusions: This study helps advance the understanding of how diverse user groups perceive and discuss widely marketed drugs like semaglutide. Although we observe a general negativity, there are nuanced differences among the subpopulations. Our results offer valuable implications for health communication strategies and pharmacovigilance.

(Online J Public Health Inform 2026;18:e80660) doi:10.2196/80660

KEYWORDS

semaglutide; public health; social media; user experience; sentiment analysis; topic modeling

XSL•FO
RenderX

Introduction

Background

Semaglutide, also known by brand names such as Ozempic and Wegovy, has surged in popularity in recent years [1]. Originally developed as a diabetes medication, semaglutide has recently shown effectiveness in off-label use for weight-loss treatment [2]. Semaglutide was the fourth-highest drug expenditure in the United States in 2021, with US $10.8 billion spent on the drug [3]. Social media platforms like X (formerly Twitter) have become key venues for the public to share experiences and express opinions about semaglutide [4]. Celebrities or public figures, such as Elon Musk, have shared personal weight-loss stories and endorsed the drug, further amplifying conversations [1,5]. The widespread advertising, endorsements by high-profile figures, and increased consumer interest have made semaglutide a trending topic in medications [6]. As such, social media offers a unique lens to examine how it is perceived and discussed, shedding light on public sentiment, misconceptions, and concerns [4,7].

Understanding public perceptions is essential, as user experiences significantly influence the evaluation of pharmaceutical drugs' effectiveness [8,9]. Positive experiences not only enhance user satisfaction but also contribute to improved adherence and overall well-being [10]. Mining social media data allows policymakers and pharmaceutical providers to tap into a vast repository of real-time data pertaining to user experiences [11]. In addition, by analyzing user-generated content, researchers can uncover nuanced insights into the concerns, preferences, and challenges faced by specific subpopulations by gender, location, or other demographic attributes. This granular analysis provides an opportunity to identify unmet needs, tailor interventions, and ensure more equitable health care outcomes.

In addition to obtaining information from social media, recent research has focused on applying natural language processing (NLP) techniques due to their ability to quickly process large-scale and unstructured information [12,13]. NLP has been leveraged to extract chemical-disease relations [14], build health knowledge graphs [15], and ease the process of documentation in electronic health records [16] (which often use unstructured and nonstandardized formats) [17]. Text analytics approaches such as named entity recognition, topic modeling, and sentiment analysis have been applied within the health context [18-20].

Although mining drug-related user experiences on social media has been widely explored [21,22], few studies have focused specifically on semaglutide-related discourse. Our study combines large-scale sentiment analysis and topic modeling with user subgroup analysis, offering a granular view of public engagement with semaglutide. Prior work has primarily leveraged social media to identify adverse reactions to semaglutide that were not detected during clinical trials [7,23]. The study by Alvarez-Mon [4] includes a manual analysis of 2045 posts to determine user interests, beliefs, and experiences pertaining to semaglutide and other antiobesity drugs. However, the public discourse including the sentiments and prevalent topics within specific user groups has been underexplored. To uncover patterns in how different user subpopulations experience and discuss semaglutide on X, we investigate the following research questions (RQs).

- RQ1 (Sentiment Analysis): What underlying factors explain how sentiment toward semaglutide evolves over time and across different user subpopulations, and what insights can be drawn about public concerns and motivations?
- RQ2 (Topic Modeling): Which topics of discussion are most prevalent in positive and negative semaglutide-related posts across different user subpopulations, how do their prevalence patterns change over time, and what do these patterns reveal about group-specific attitudes, priorities, and health communication needs?

The first research question aims to explore differences in engagement patterns and sentiment expressions across user subpopulations and over time. The second question aims to identify the various discussion topics emphasized by distinct user subpopulations, exploring the prevalence of these topics among the subpopulations. Addressing these two research questions, our study provides a comprehensive discourse analysis across user subgroups, along with an identification of external events that influence the evolution of these patterns. This insight into the real-world user experience is crucial for tailoring public health communication and improving medicine support strategies for diverse communities.

Related Work

Exploring User Experiences via Social Media

Crowdsourcing, originally defined as the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call, has revolutionized how researchers gather and analyze public opinion [24]. Similar methodological approaches have been applied in other domains to understand public perceptions of urban accessibility and inclusion through crowdsourced online reviews [25]. While conventional methods like polls and surveys remain valuable, crowdsourcing through social media platforms enables researchers to collect and analyze large-scale, near-real-time data about user experiences and perspectives [26].

In the health care domain, crowdsourcing via social media has become particularly valuable for understanding public opinions about medical treatments and pharmaceutical drugs [27,28]. First, social media data have shown invaluable potential in pharmacovigilance due to the tendency of users to share their opinions or experiences such as adverse drug reactions [29]. Researchers have leveraged social media data to track topic trends [30], estimate disease prevalence [31], and analyze public response to health policies [32]. Second, crowdsourcing has proven effectiveness in capturing user experiences that might not be readily available through traditional clinical studies or surveys [33]. In particular, social media platforms provide researchers with access to diverse user populations and their real-world experiences with pharmaceutical drugs, such as off-label use [9] and adverse reactions [34].

XSL•FO
RenderX

Analyzing User Experiences via NLP Techniques

Our work makes use of sentiment analysis and topic modeling to uncover patterns in user experiences with semaglutide. Sentiment analysis, also called opinion mining, is a branch of NLP that focuses on classifying people's opinions into positive, negative, or neutral associated with data [35]. Research in this domain spans various levels of granularity, from assigning a single sentiment to an entire document or individual sentences to analyzing distinct aspects linked to specific entities [36]. After the COVID-19 pandemic, there has been increasing interest in using sentiment analysis to evaluate the attitudes, perceptions, and emotions expressed by social media users [37-39]. Numerous studies have focused on platforms such as X, Reddit, and Facebook, which have become prominent spaces for sharing public opinions related to COVID-19 [40].

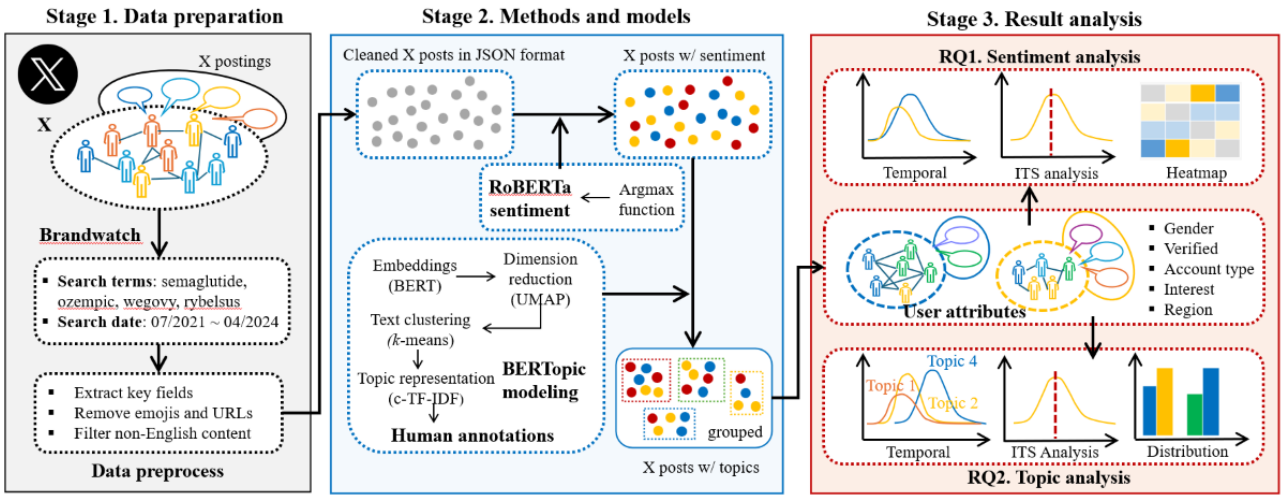
Topic modeling techniques, such as latent Dirichlet allocation [41] and BERTopic [42], seek to discover the key themes present in a corpus of documents [43]. Topic models can summarize large datasets by capturing the topics (ie, the major discourse) that appear most commonly in the text. Numerous works have used topic models to study health-related discussions on social media. For example, Asghari [44] identifies trending topics pertaining to health care on X. Topic analysis has provided insights into news reports surrounding COVID-19 [45] and public opinion regarding blood donation [46]. Another study trains an aspect-based topic model to characterize the health topics and then estimates the prevalence of influenza and allergies over time by observing the number of mentions of each topic during different time periods [47]. In addition, prior work has analyzed semaglutide-related discussions on Reddit via topic modeling [48-50]. Our study focuses on exploring how public discourse and sentiment differ across user subpopulations. Analyzing specific subpopulations enables a nuanced understanding of user concerns, such as accessibility, side effects, and insurance coverage, guiding targeted strategies for addressing subpopulation-specific needs.

Methods

Overview

Figure 1 provides an overview of our research design. We compile a dataset consisting of semaglutide-related X posts (stage 1) and perform sentiment analysis and topic modeling on these posts (stage 2). We analyze the results to address our two research questions (stage 3).

Figure 1. Overview of the designed framework to implement the research. BERT: bidirectional encoder representations from transformers; c-TF-IDF: class-based term frequency-inverse document frequency; ITS: interrupted time series; RQ: research question; UMAP: uniform manifold approximation and projection.



Data Preparation and User Attributes

We use Brandwatch [51], a social media analytics platform, to collect data from X. Brandwatch uses the X application programming interface to obtain posts from prior periods, offering a representative sample of X's entire dataset. Our data collection targets X posts posted between July 1, 2021—when the US Food and Drug Administration (FDA) approved semaglutide for chronic weight management [52]—and April 30, 2024. This nearly 3-year timeframe enables our longitudinal discourse analysis surrounding semaglutide within different user communities.

To collect the data, we use several key search terms, including "semaglutide" and its branded names "Ozempic," "Wegovy," and "Rybelsus." We chose these search terms to ensure comprehensive coverage of discussions related to semaglutide and its marketed variants. To maintain consistency, we limit the dataset to English-language X posts containing these terms. The final dataset consists of 859,751 posts, including original posts, replies, reposts (retweets), and quotes.

The user attributes we study are gender, US region, interests, account type, and verification status. These attributes are inferred by the data provider, Brandwatch. According to their documentation, the methodology for this inference varies by attribute; for instance, location is primarily determined from explicit, user-provided information in their X public profiles, whereas attributes like gender and interests are classified using machine learning models that analyze public data such as first names and biographical text [53]. We acknowledge that the specific algorithms used for this demographic inference are proprietary to Brandwatch, and as such, the details of their methods are not publicly available [53].

XSL•FO
RenderX

We choose these attributes for their relevance in capturing diverse user perspectives and behaviors. For example, gender and region can reveal variations in health care access and cultural attitudes [54,55] and have been identified as factors associated with semaglutide initiation [56]. Account type and verification status help differentiate individual users from organizations. Interests reflect personal activities. It should be noted that users may have multiple interests or none, placing them in zero or more interest-based subpopulations.

Since individual users may post multiple posts, we measure the size of each subpopulation by the number of users and by the number of posts. When grouping each post by its respective user, we assign aggregated values for attributes (gender, region, interests, account type, and verification) using the most frequently observed values across that user's posts. This aggregation ensures that each user is represented by a single record while preserving their attributes. The dataset contains 436,551 unique users. Tables S1 and S2 in Multimedia Appendix 1 summarize the user subpopulations.

Ethical Considerations

The University of South Florida Institutional Review Board reviewed this study and determined it to be exempt (STUDY009222). To remain compliant with X's policies, we adhere to all application programming interface rate limits and do not collect deleted posts. All the code used to generate our results is available from our GitHub repository [57]. To protect user privacy, our full dataset containing identifiers such as usernames is not available for public access [58]. An anonymized version can be accessed from Multimedia Appendix 1 or requested from the corresponding author. Lastly, we follow standard guidelines [59] to mitigate potential harms from sensitive content contained in our dataset (eg, suicide mentions).

RoBERTa Sentiment Analysis

To classify the sentiment of X posts, we use the *cardiffnlp/twitter-roberta-base-sentiment-latest* model from Hugging Face [60], which is pretrained on X data and widely recognized for its state-of-the-art performance in sentiment classification tasks [61], particularly in handling X posts [62]. We apply the RoBERTa (Robustly Optimized BERT Pretraining Approach) model to each post, obtaining a set of values representing the likelihood of the post having negative, neutral, or positive sentiment. The model outputs these values in a list format, with the elements at indices 0, 1, and 2 corresponding to the negative, neutral, and positive probabilities, respectively [60]. The final sentiment label is determined by the *argmax* function; no additional calibration or thresholding is performed on the sentiment scores. After that, we assign each post a sentiment label, that is, –1 for negative, 0 for neutral, and 1 for positive. As a result, of the 859,751 total posts, 116,091, 429,074, and 314,586 are classified as positive, neutral, and negative sentiment, respectively.

We calculate mean sentiment scores within each of the user subpopulations (gender, US region, account type, verification status, and interests). To measure sentiment per user, we first group each post by its respective user. We then calculate the mean sentiment score for each user by averaging their per-post sentiment scores, resulting in a continuous value between –1 and 1. This grouping allows us to analyze the aggregated averages of user-level sentiments across the subpopulations. Since the volume of posts varies by subpopulation (eg, individual user accounts create 1.9 posts on average, compared to 4.4 from organizational users), we also calculate average sentiment per post. In addition, to assess the robustness of our sentiment findings, we conduct a repost-excluding sensitivity analysis, in which we measure average sentiment within each subpopulation while excluding reposts (resulting in 411,747 posts for sensitivity check).

Interrupted Time-Series Regression

To examine how sentiment evolves over time, we group the dataset and calculate the average sentiment bimonthly spanning from July 2021 to April 2024. This procedure yields a time-series of average sentiment scores. Our longitudinal analysis is performed per post, as a user can publish multiple posts at different times. To determine whether the observed sentiment shifts exceed baseline trends, we apply an interrupted time-series (ITS) regression. Using bimonthly sentiment averages, we model (1) the baseline time trend, (2) an immediate level change at the intervention point, and (3) a slope change after the event. This approach allows us to separate long-term temporal patterns from abrupt discontinuities. Given the large dataset and multiple comparisons, we treat results as exploratory and emphasize effect sizes and CIs over strict hypothesis testing. Our ITS regression is modeled as follows:

where Y_t represents the mean sentiment score during the bimonthly period, $Time_t$ is the continuous time index, $Event_t$ is a binary indicator coded 0 before the intervention and 1 afterward, and $Time_tEvent_t$ captures the postevent slope. Here, β_0 estimates the baseline level at the beginning of the series, β_1 captures the pre-event trend, β_2 reflects the immediate level change at the intervention, and β_3 represents the postevent slope change.

BERTopic Modeling

After performing sentiment analysis, we use the BERTopic model [42] to discover the commonly discussed topics in the dataset. We divide the dataset into posts with a positive RoBERTa sentiment label and posts with a negative sentiment label; posts with neutral sentiment are excluded to focus on identifying the topics that contribute to positivity and negativity. We perform topic modeling for the positive and negative tweets separately. To create the positive and negative document corpora from our dataset, we first clean the text of each post. This cleaning process involves steps such as removing emojis, punctuation, and stop words, normalizing whitespace, and converting all text to lowercase. Cleaning the text is an important step due to the unstructured nature of social media posts [63]. Note that these cleaning steps are not performed prior to obtaining the sentiment of each post, as they may have affected the sentiment results. For example, emojis [64] and punctuation [65] can impact sentiment scores. Our topic modeling exercise focuses solely on the themes present in the text, as opposed to

XSL•FO
RenderX

the sentiment of the text. After cleaning each post's text, we remove duplicated reposts to avoid situations in which distinct reposts of a given post are mapped to different topics. The positive and negative document corpora are lists consisting of the snippets of each of the positive and negative cleaned X posts, respectively.

After performing the cleaning steps, we initialize a BERTopic model with default values for all hyperparameters. We run the BERTopic model on our document corpora to generate a list of topics and then extract the sentence/document embeddings for those topics. We use *k*-means clustering [66] with the Elbow method to determine the optimal number of topics. Figures S1 and S2 in Multimedia Appendix 1 show the *k*-means clustering results for the positive and negative document corpora, respectively. We then run the BERTopic model on our document corpora using 100 clusters.

BERTopic outputs a topic representation and a document representation. In our case, the documents are cleaned X posts. For each topic, the topic representation lists its representative keywords, representative documents, and document count. The BERTopic document representation maps each document in the corpus to its topic number. After reviewing the 100 positive and 100 negative topics produced by BERTopic, we observe that many of the topics shared similar topics. We therefore manually annotate the 200 topics into 10 umbrella topics. These topic groupings consist of all the topics that share a common theme. For instance, umbrella topic 0 consists of topics relevant to weight loss.

We perform the manual annotation by reviewing the representative keywords and documents. For example, one of the most common positive topics is represented by the following list of keywords: [*semaglutide*, *semaglutides*, *weightloss*, *diet*, *appetite*, *medication*, *eat*, *treatment*, *craving*, *fda*]. A representative document (X post after cleaning) for this topic is: "ready lose weight gain confidence say hello semaglutide gamechanging prescription medication help achieve significant weight loss with semaglutide take control craving appetite finally reach weight loss goal." These keywords and the document are associated with themes of weight loss, so we map this topic to umbrella topic 0 (weight loss). As another example, one of the most common negative topics is represented by these keywords: [*nausea*, *diarrhea*, *vomiting*, *nauseous*, *constipation*, *vomit*, *nauseate*, *constipate*, *diarrhoea*, *stomach*]. One of the representative documents for this topic is: "nausea diarrhea stomach abdominal pain vomiting constipation side affect ozempic." This topic pertains to adverse reactions experienced after taking semaglutide; therefore, we map it to umbrella topic 8 (acute harm/adverse drug reactions). The mapping of all 200 initial topics to the 10 umbrella topics is available in Multimedia Appendix 2. Additionally, to verify that the umbrella topic shares do not change significantly under different cluster counts, we conduct stability checks using 50 and 25 clusters. The results are described in the Umbrella Topic Stability Checks section in Multimedia Appendix 1.

Two of the authors (PM and GL) independently map the initial 200 topics to the 10 umbrella topics. We calculate the intercoder agreement between the annotators using Krippendorff α [67], as described by the following equation, where D_o is the observed disagreement between the annotators and D_e is the expected random disagreement:

The intercoder agreement is 0.806, indicating a satisfactory level of agreement between the annotators. Each X post is mapped to one of the clustered 200 topics using BERTopic, and each of the clustered 200 topics is mapped to one of the 10 umbrella topics based on manual annotation. We can therefore map each post to its umbrella topic. Table 1 shows the 10 umbrella topics, examples of their representative keywords, and the number of posts mapped to each topic. In addition, Table S3 in Multimedia Appendix 1 provides example representative documents (posts) for the 10 umbrella topics. During the annotation, the topics are mapped to umbrella topic T9 ("other") if the topic does not clearly match any of the other 9 umbrella topics. We observe various subtopics that appear within T9, such as drug marketing/news, other nonsemaglutide drugs, and health/beauty. With the labeled umbrella topics, we group the data by user attributes to discover the prevalence of each umbrella topic among user subpopulations.

XSL•FO
RenderX

Table 1. Number, name, and representative keywords for each of the 10 umbrella topics.

Topic no.	Topic name	Posts, n	Example representative keywords
T0	Weight loss	50,542	exercise, workout, eat, diet, appetite, skinny, obesity, craving
T1	Celebrities/politicians	40,120	nikkifried, erikajayne, oliviawilde, rhianna, tuckercarlson, oprah, elonmusk, trump
T2	Obtaining the drug	22,724	prescription, medication, walgreens, walmart, coupon, insurance, coverage, supply, ordered, appointment, affordable, shot, injection
T3	Drug indicators	16,948	diabetes, inflammation, treatment
T4	Drug authorities	23,783	pharma, novo, nordisk, doctor, physician, fda, goldman, economy, gdp, market
T5	General and profane negativity	9488	[swear words]
T6	Suicide risk	1486	suicide, autopsy, death, overdose
T7	Chronic harm	3075	addiction, cancer, tumor, alocepia, hair, hairline, dialysis
T8	Acute harm/adverse drug reactions	10,208	nausea, diarrhea, constipation, pain, effect
T9	Other	29,475	[anything that does not fit in with the other topics]

Results

RQ1 (Sentiment Across User Subpopulations)

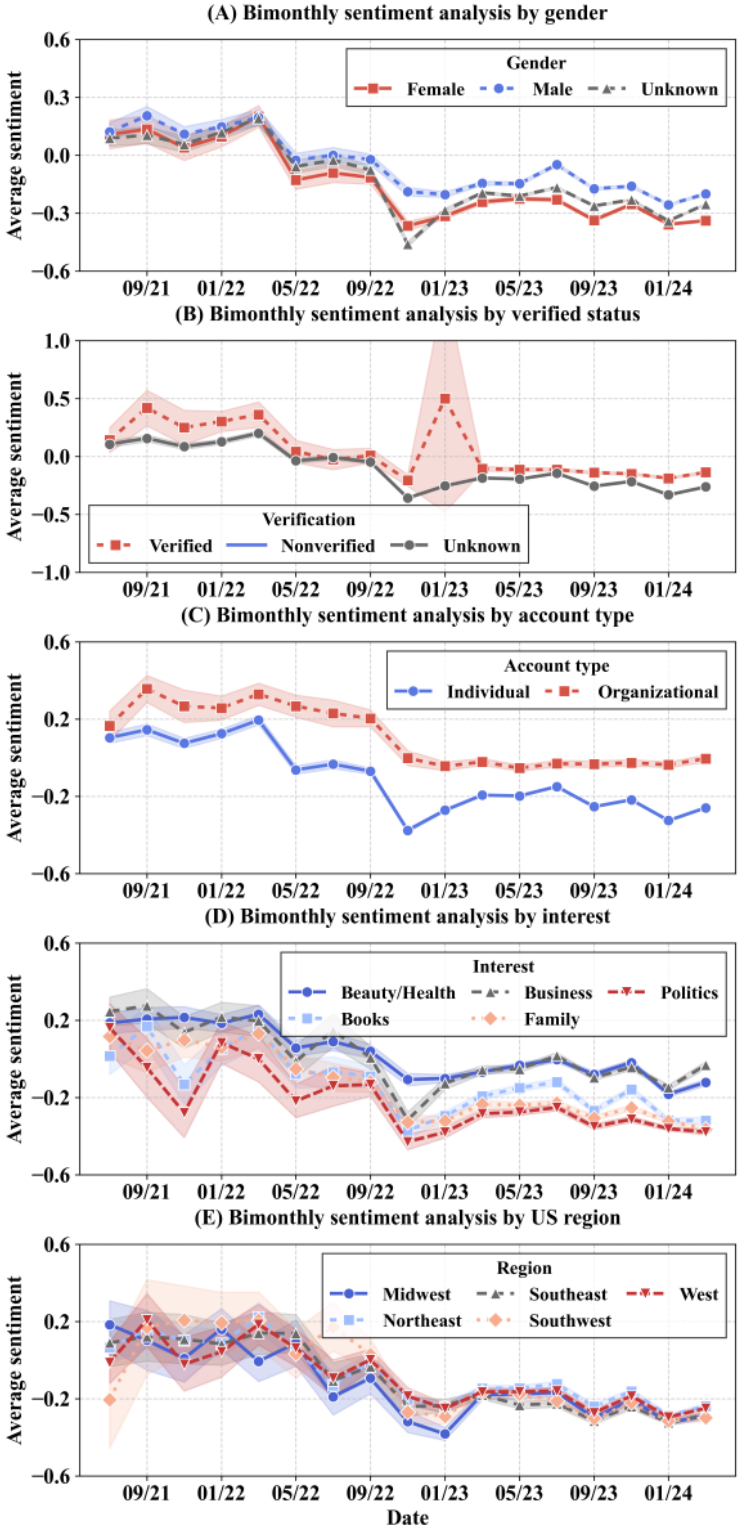
Addressing RQ1, we find that the overall sentiment toward semaglutide during the study period (July 2021 to April 2024) is slightly negative, with a mean sentiment score of –0.24 (SD 0.669) across all posts and –0.28 (SD 0.605) across all users. Within all user subpopulations, we observe a decline in sentiment over time, but the trend varies across different subpopulations, as discussed in the following subsections.

Overall Sentiment Declines

Our longitudinal sentiment analysis results are displayed in Figure 2. To illustrate uncertainty, we include 95% CIs for key categorical attributes as shaded bands. The temporal progression of sentiment can be divided into four phases: (1) initial positive sentiment across categories (2021 to mid-2022), (2) universal decline (November 2022 to January 2023), (3) variable recovery rates through 2023, and (4) eventual stabilization at slightly negative levels by early 2024. Three notable periods of universal decline are present: the largest from November 2022 to January 2023, the second from September to November 2023, and the final from January to March 2024. Our repost-excluding sensitivity analysis yields similar trends, supporting the robustness of these temporal trends.

XSL•FO
RenderX

Figure 2. Time series plots showing bimonthly sentiment analysis categorized by gender (A), verified status (B), account type (C), interest (D), and region (E). To improve readability, we limit the interest visualization to the top 5 most popular of the 21 user interests. Shaded areas represent 95% CIs.



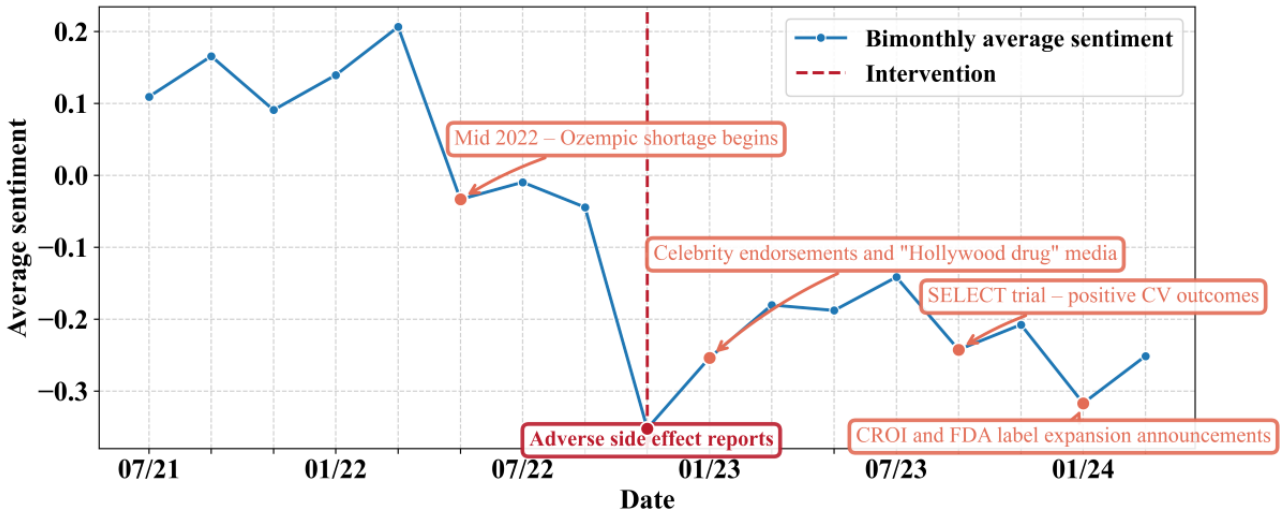
Connection With External Events

The sentiment changes are associated with external events (some key events are highlighted in Figure 3). The first clear sentiment decline starting in mid-2022 coincides with the national shortage of glucagon-like peptide 1 medications, which led to increased reliance on compounded semaglutide formulations and raised concerns about access, safety, and regulatory oversight [68]. During this same period, the FDA approved Wegovy for adolescent patients, intensifying public debate around expanded clinical use [69]. Moreover, there is a pronounced decline in sentiment within all user subpopulations during November 2022–January 2023, which coincides with reports of adverse gastrointestinal reaction, suggesting a public concern about potential side effects [70]. Following this decline, a temporary

XSL•FO
RenderX

spike in positive sentiment, particularly among verified users, emerges between January 2023 and March 2023. However, the wide CI for verified users indicates a high standard error likely resulting from the limited number of verified-user posts within that bimonthly window. This increase in sentiment coincides with heightened media attention and celebrity endorsements of semaglutide as a "Hollywood weight loss drug" [71], suggesting that promotional activity and influencer-driven narratives could temporarily reverse prevailing sentiment patterns.

Figure 3. Interrupted time series analysis of average sentiment over time with external events. The dashed red line indicates the start of the intervention window. CROI: Conference on Retroviruses and Opportunistic Infections; CV: cardiovascular; FDA: US Food and Drug Administration; SELECT: Semaglutide Effects on Cardiovascular Outcomes in People With Overweight or Obesity.



Following the second dip between September and November 2023, sentiment increases across all user attributes. This recovery aligns with the release of positive cardiovascular outcomes from the SELECT (Semaglutide Effects on Cardiovascular Outcomes in People With Overweight or Obesity) trial, which was presented at the American Heart Association Scientific Sessions and published in the *New England Journal of Medicine* in November 2023 [72]. The results show that Wegovy (semaglutide 2.4 mg) can significantly reduce major cardiovascular events in adults with overweight or obesity and established cardiovascular disease, potentially improving public perceptions of the drug's efficacy and safety. Following the third dip in sentiment between January and March 2024, we observe a rebound across nearly all subpopulations. This recovery coincides with two major announcements in early March 2024: (1) a National Institutes of Health–sponsored study presented at the 2024 Conference on Retroviruses and Opportunistic Infections showed semaglutide significantly reduces liver fat in people with HIV and MASLD (metabolic dysfunction-associated steatotic liver disease) [73] and (2) the FDA approved a label expansion for Wegovy to include cardiovascular risk reduction based on long-term SELECT trial data [74]. These developments likely contributed to renewed optimism about semaglutide's broader therapeutic value.

However, the sentiment patterns among user groups vary. For example, we observe that organizational accounts and verified users have consistently more positive sentiment scores than individual accounts and unverified users. Users who are interested in "business" exhibit moderate sentiment scores, reflecting the professional nature of corporate communications. Users interested in "politics" exhibit the most negative and volatile sentiment, whereas the "beauty/health" category maintains the most positive sentiment among all groups throughout the timeline. Users with an interest in "books" maintained the most stable sentiment pattern. These patterns highlight how different user communities process health-related information through their respective contextual frameworks, with users valuing family and personal health showing the highest sentiment variation.

The Impact of Semaglutide Shortage and Side Effect Reports

In addition, we observe a significant downturn in sentiment across all user subpopulations during November 2022–January 2023. This universal decline aligns temporally with reports of semaglutide shortages [75-77] and reports of adverse drug reactions [70], suggesting a public concern regarding the drug's availability and potential side effects. We apply an ITS regression to assess the impact of these shortages and determine whether the sharp decline in sentiment exceeds baseline trends, as displayed in Table 2. The baseline sentiment is significantly positive (β_0=0.172, *P*=.003), with a modest but significant negative trend prior to the event (β_1=−0.027, *P*=.03). At the intervention point, an immediate and statistically significant drop occurs (β_2=–0.430, *P*=.004). While the postintervention slope shows a slight positive trend (β_3=0.029, *P*=.07), this effect does not reach conventional significance thresholds.

XSL•FO
RenderX

Table 2. Interrupted time-series regression of average sentiment. The intervention point is set at November 2022–January 2023; post time indicates the slope change thereafter.

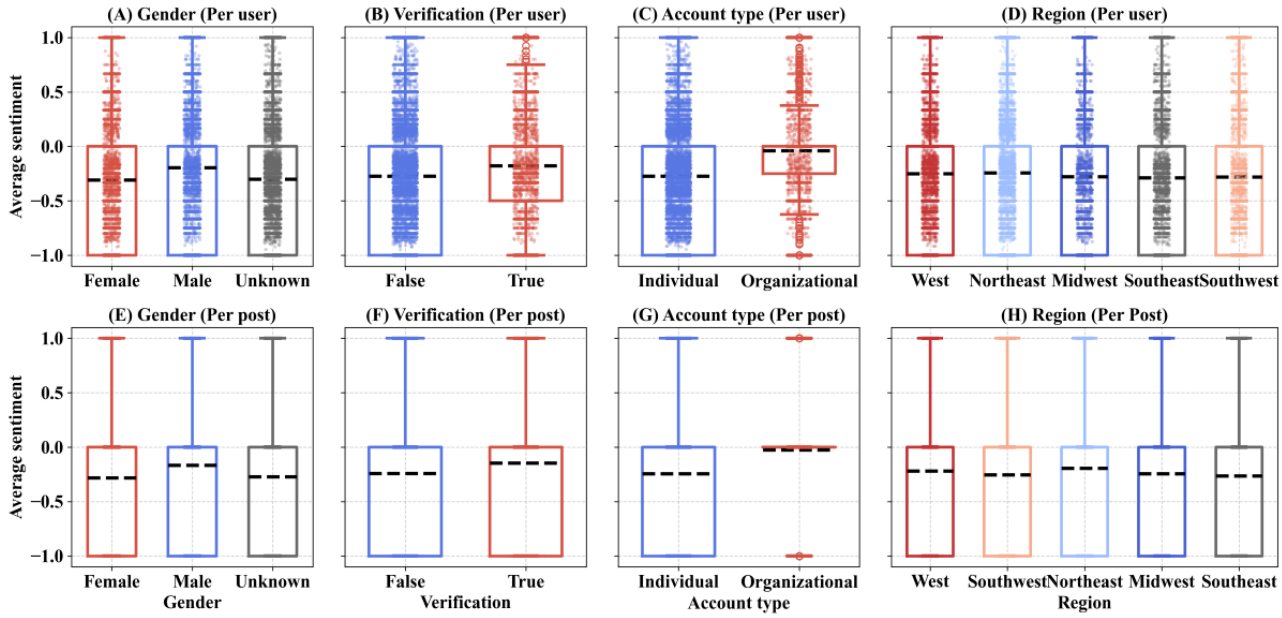
Variable	β coefficient (95% CI)	SE	*P* value
Constant	0.172 (0.070 to 0.274)	0.047	.003
Time	–0.027 (–0.051 to –0.003)	0.011	.03
Event	–0.430 (–0.700 to 0.160)	0.125	.004
Post time	0.029 (–0.00 to 0.06)	0.015	.07

Figure 3 visualizes the ITS regression, displaying observed mean sentiment across bimonthly intervals. A vertical dashed line marks the November 2022 intervention, illustrating changes in level and slope. As shown in Figure 3, sentiment remains relatively stable and slightly positive throughout 2021 and most of 2022. A clear decline emerges in the November 2022-January 2023 period, after which sentiment consistently remains below zero, indicating a sustained downturn in public discourse. The ITS results suggest that the intervention period coincides with a significant immediate downturn in sentiment, followed by a slight, nonsignificant recovery trend thereafter.

Sentiment Patterns Across Subpopulations

We observe that the per-user and per-post sentiment scores are similar; therefore, to avoid duplicated explanation and focus on user experience, we solely discuss per-user scores in this section (per-post scores and repost-excluded scores are available in Table S5 in Multimedia Appendix 1). The sentiment distribution among user subpopulations varies significantly, as illustrated in Figure 4. In addition, Table 3 presents 95% CIs for the estimated mean sentiment differences across key user groups. All comparisons yield statistically significant results ($P<.001$), providing strong evidence that the observed differences are not due to random variation.

Figure 4. Sentiment patterns across user subpopulations, measured per user and per post (the sentiment of a post can only be the value from –1, 0, and 1).



XSL•FO
RenderX

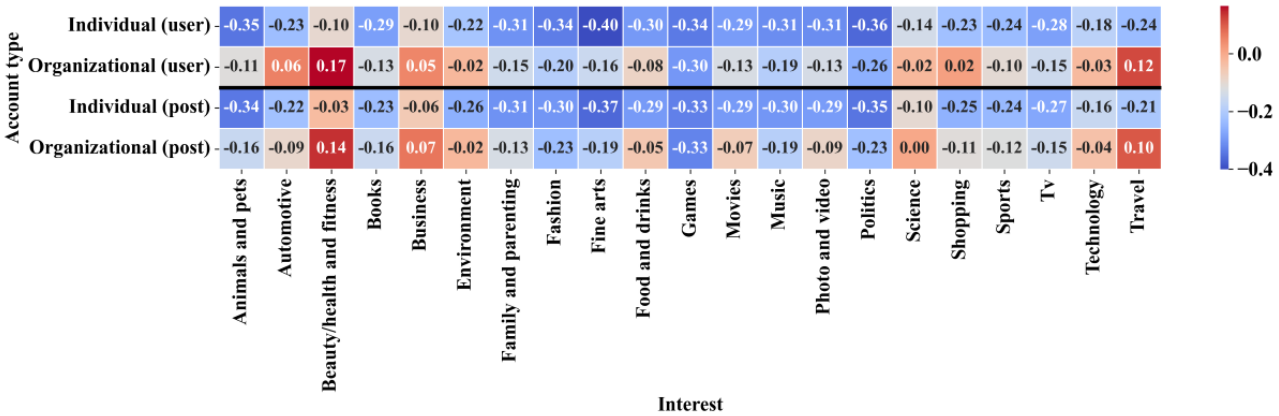
Table 3. Estimated differences in sentiment across key subgroups, with 95% CIs.

Comparison and levels	Mean 1	Mean 2	Mean difference (95% CI)	*P* value
Male vs female				
Per post	–0.17	–0.29	0.12 (0.121-0.129)	<.001
Per user	–0.20	–0.32	0.12 (0.117-0.128)	<.001
Verified vs nonverified				
Per post	–0.14	–0.24	0.10 (0.092-0.102)	<.001
Per user	–0.17	–0.28	0.11 (0.101-0.116)	<.001
Organizational vs individual				
Per post	–0.01	–0.24	0.23 (0.222-0.235)	<.001
Per user	–0.04	–0.28	0.24 (0.232-0.256)	<.001

Based on Figure 4, male users exhibit more positive sentiment (averaged sentiment –0.20) than female users (averaged sentiment –0.32). Verified users, typically public figures or organizations with confirmed identities, expressed less negative sentiment (averaged sentiment –0.17) toward semaglutide than nonverified users (averaged sentiment –0.28). This contrast highlights how identity and accountability influence sentiment expression online. Organizational accounts expressed less negative sentiment compared to individual users. These findings suggest that organizations tend to frame their discussions about semaglutide in a more positive manner, possibly due to pharmaceutical marketing, professional communication standards, or endorsement practices. Within the United States, regional variations in sentiment toward semaglutide are evident in the analysis. Users in the Southeast region express the most negativity (averaged sentiment –0.29), while users from the Northeast are less negative.

Lastly, Figure 5 presents the average sentiment scores for each combination of account type and user interest, measured per user and per post. Overall, individual users tend to express more negative sentiment than organizational accounts. For example, in the "travel" category, individual sentiment is clearly negative (–0.24), while organizational sentiment is positive (0.12), yielding one of the largest absolute gaps between the two account types. This figure highlights that user sentiment varies widely across domains of interest. Health-related categories, particularly those tied to "beauty/health and fitness," generate a higher level of positivity, while "business" discussions are closer to neutral, reflecting a more corporate and less personal orientation. These findings underscore the importance of considering both account type and user interest when analyzing public sentiment.

Figure 5. Combined heatmap of mean sentiment scores by account type and interest.



RQ2 (Topic Results Across User Subpopulations)

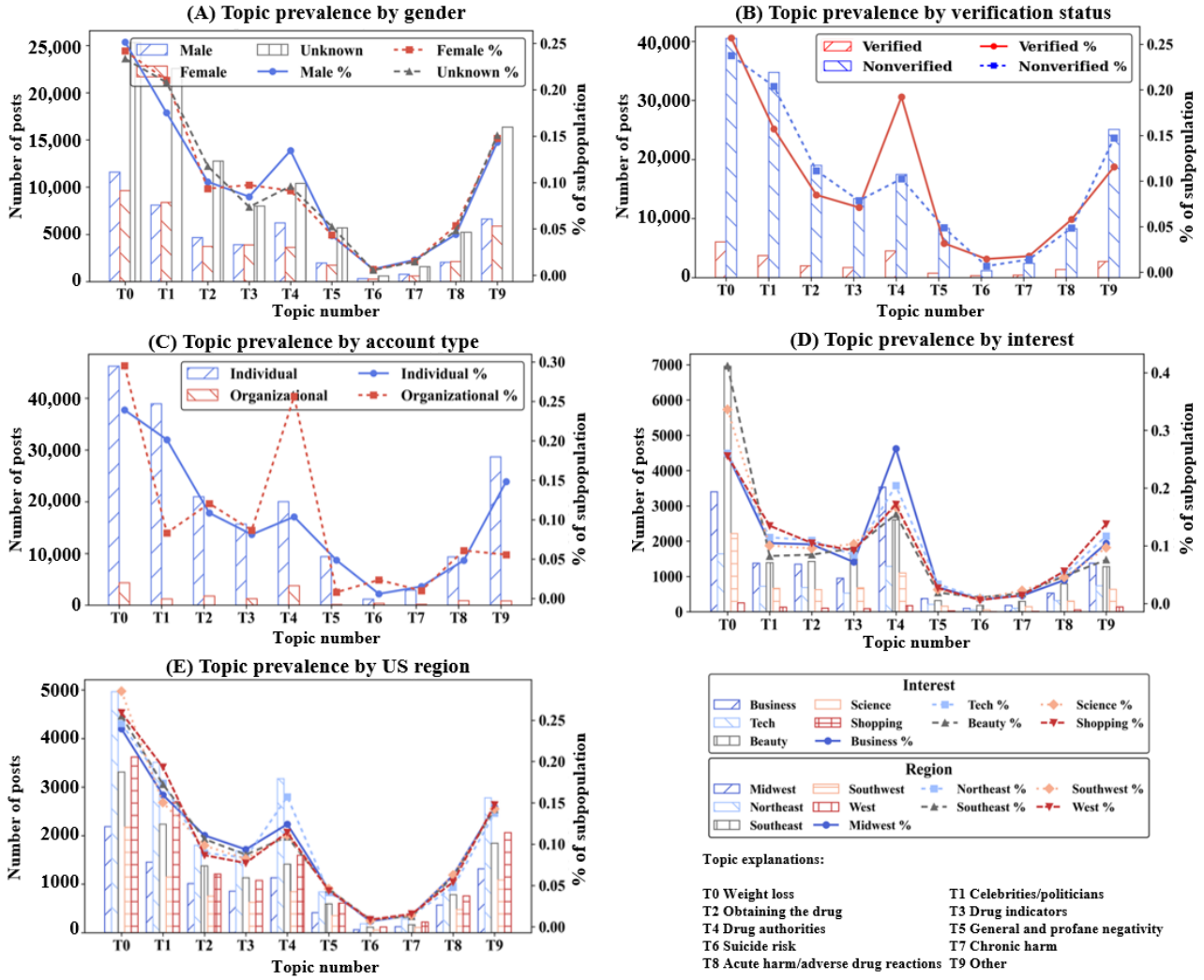
For the topic analysis, we focus on original posts by successfully assigning umbrella topics to 207,849 out of 859,751 posts. Not all posts in the dataset receive topic assignments because we exclude duplicated reposts (to avoid counting the same content multiple times) and neutral-sentiment posts from the topic modeling process. We conduct topic modeling analysis by post rather than aggregating by user, which preserves the complete topic distribution for prolific users.

Topic Prevalence Across Subpopulations

Figure 6 shows the prevalence of each topic within each user subpopulation, highlighting the variations in each topic's prevalence across subpopulations. The most common topic among subpopulations was T0 (weight loss). However, the popularity of the other topics is less consistent across the subpopulations. We present a descriptive analysis of the topic prevalence results in the following paragraphs; CIs and effect sizes (calculated via Cramér *V* [78]) for these results are given in Table S7 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 6. Prevalence of each topic within each user subpopulation. The bars and left axes measure the number of posts pertaining to each topic, and the lines and right axes measure the percentage of each subpopulation's posts that pertain to each topic.



T1 (celebrities/politicians) is noticeably more popular among female users compared to male users, comprising 21% (8368/39,786) of posts from female users and 17.5% (8087/46,133) of posts from male users. Although there are more overall posts originating from male users, female users post more T1 posts. In addition, 13.4% (6200/46,133) of male user posts pertained to T4 (drug authorities), compared to 9.1% (3630/39,786) of posts from female users.

Verified users are less likely to post profane posts, with T5 (general and profane negativity) comprising 3.2% (748/23,549) of their posts, compared to 4.9% (8318/170,279) of posts from unverified users. However, verified users are about twice as likely to post about T6 (suicide risk) than unverified users. They are also more likely to create posts pertaining to T4 (drug authorities); 19.2% (4525/23,549) of verified user posts belong to T4, compared to 10.3% (17,460/170,279) from unverified users.

Examining the most prevalent topics among individual accounts and organizational accounts, the most striking difference is the very low number of profane posts within the organizational account subpopulation. T5 comprises just 0.8% (117/14,711) of organizational posts, compared to 4.9% (9371/193,138) of posts from individual users. As companies and organizations likely do not want to damage their reputation by posting profane content, this result is in line with our expectations. On the other hand, organizations are about 4 times more likely than individuals to create posts pertaining to T6 (suicide risk), perhaps due to medical organizations posting warnings about potential side effects of semaglutide. T4 is far more common among organizational accounts, with 25.7% (3774/14,711) of organizational posts belonging to T4 compared to 10.4% (20,009/193,138) from individual users.

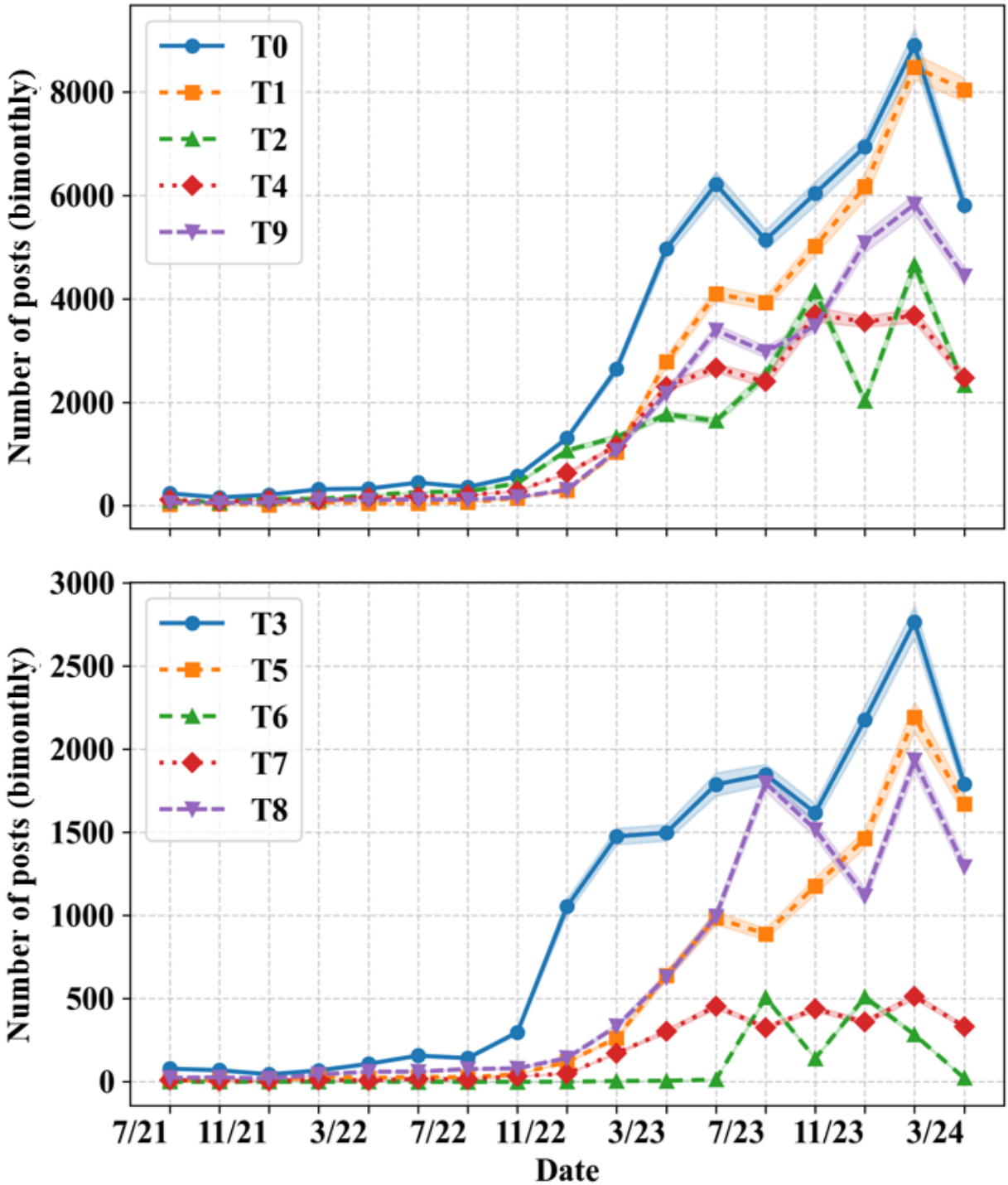
Dividing the users by interest reveals several differences in topics of discussion. Notably, the subpopulation consisting of users interested in "business" is the only subpopulation in which T0 is not the most prevalent topic. T4 is the most common topic among users interested in "business." Users in the "business" subpopulation appear to be more interested in the economic impact of semaglutide, as opposed to its usage in weight loss treatment. As expected, T0 is by far the most popular topic among users interested in "beauty/health." Lastly, the prevalence of each topic is mostly consistent across different US geographic regions. However, there are some variations; for example, T4 is noticeably more popular in the Northeast compared to other regions.

XSL•FO
RenderX

Topic Prevalence Over Time

Figure 7 shows the number of posts pertaining to each umbrella topic posted during each bimonthly period from July 2021 to April 2024. To assess the evolution of topic prevalence over time, we first present an exploratory analysis of external events that may have influenced the topic trends. All topics rose in popularity from July 2021 to April 2024. This result is consistent with the general increase in popularity of semaglutide. T3 surged in popularity during November 2022-January 2023. This trend aligns with the initial FDA approval of Wegovy for adolescents [69], which occurred on December 23, 2022. During September 2023-November 2023, T0 (weight loss) increases in prevalence, while T8 (acute harm/adverse drug reactions) and T6 (suicide risk) decline. These changes align with the release of positive cardiovascular outcomes for Wegovy from the SELECT trial [72].

Figure 7. The number of posts pertaining to each topic over time. T0: weight loss; T1: celebrities/politicians; T2: obtaining the drug; T3: drug indicators; T4: drug authorities; T5: general and profane negativity; T6: suicide risk; T7: chronic harm; T8: acute harm/adverse drug reactions; T9: other.

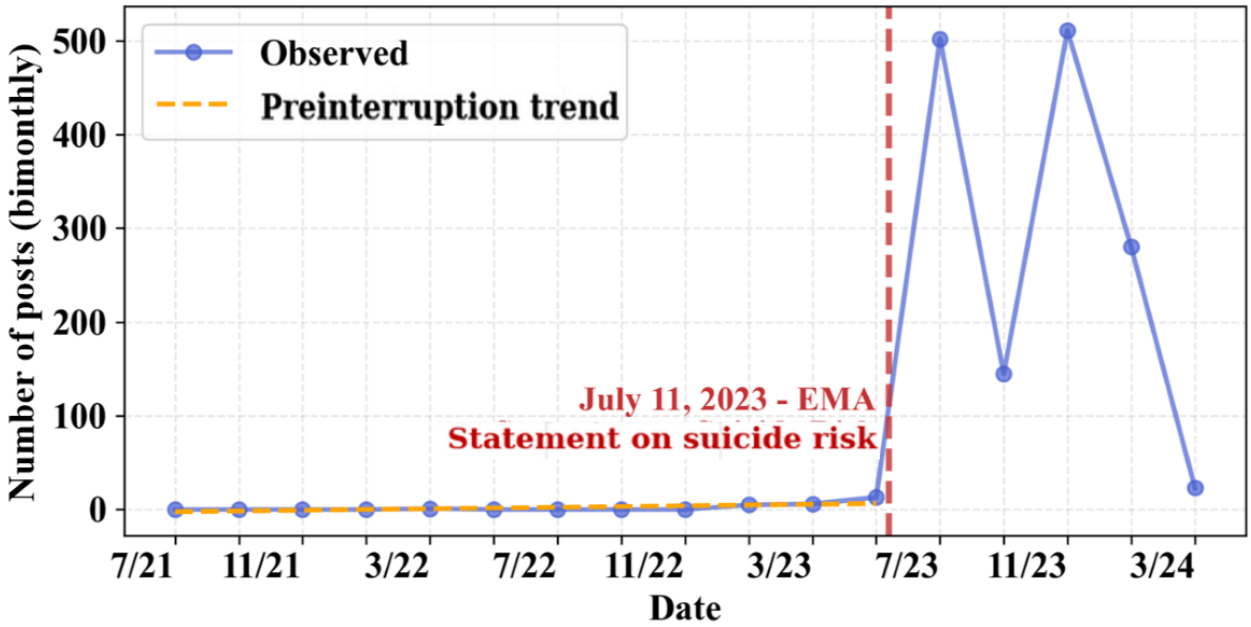


Starting in July 2023–September 2023, there is a noticeable uptick in the number of posts pertaining to T6 (suicide risk) and T8 (acute harm/adverse drug reactions). To determine whether this change exceeds baseline trends, we apply an ITS regression. The ITS results are visualized in Figure 8 (full results are available in Table S6 in Multimedia Appendix 1). A statistically

XSL•FO
RenderX

significant (β_2=449.51, P<.001) increase occurs at the intervention point, followed by a significant downward trend (β_3=–83.10, P=.02). The ITS results suggest that the intervention window coincides with a sharp increase in the prevalence of T6. The surge in the popularity of T6 may have been caused by a statement released by the European Medicines Agency on July 11, 2023, acknowledging "about 150 reports of possible cases of self-injury and suicidal thoughts" from "people using liraglutide and semaglutide medicines" [79].

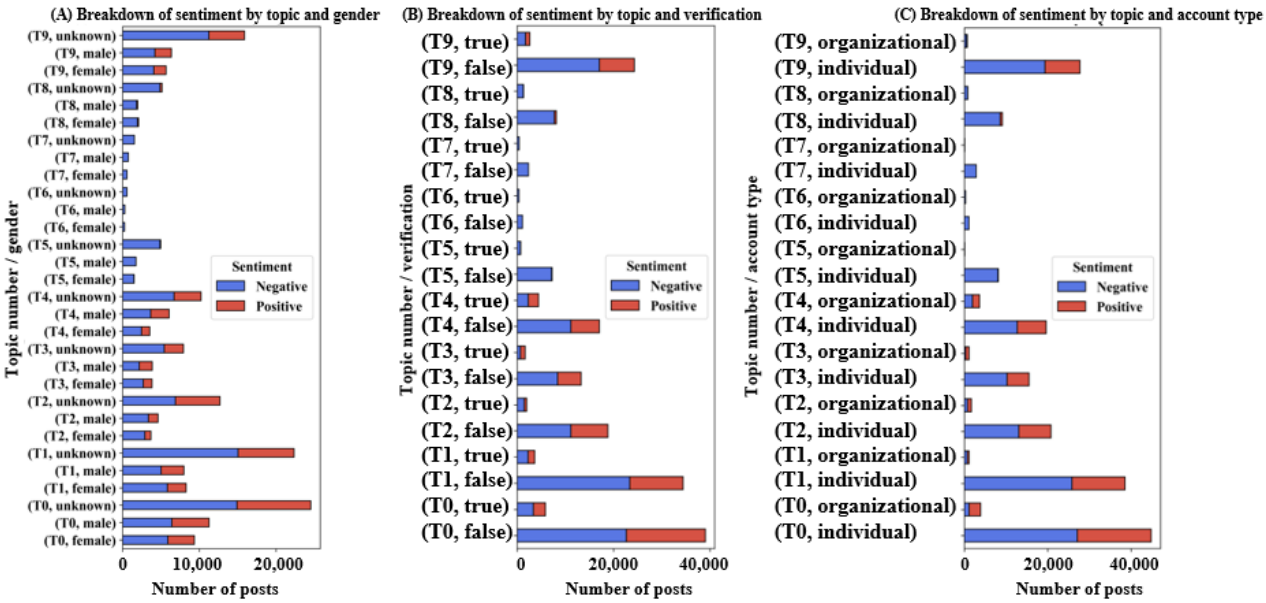
Figure 8. Interrupted time series analysis of T6 (suicide risk). The intervention window is July 2023–September 2023. EMA: European Medicines Agency.



Breakdown of Topics by Sentiment

Figure 9 shows the breakdown of sentiment scores within each topic, within the gender, verification status, and account type subpopulations. This figure connects our sentiment and topic analyses and provides insights into potential sources of user negativity and positivity. For example, topics T5 (general and profane negativity), T6 (suicide risk), T7 (chronic harm), and T8 (acute harm/adverse drug reactions) are almost entirely negative.

Figure 9. Breakdown of sentiment scores by topic number, within various user subpopulations. T0: weight loss; T1: celebrities/politicians; T2: obtaining the drug; T3: drug indicators; T4: drug authorities; T5: general and profane negativity; T6: suicide risk; T7: chronic harm; T8: acute harm/adverse drug reactions; T9: other.



The breakdown also reveals notable differences in sentiment within topics. In T4, for instance, male-identified accounts contribute a significant volume of positive posts, whereas female-identified accounts are almost exclusively negative (Figure 9A). A similar, though less pronounced, pattern is visible in T9, where male accounts again show a larger share of positive

XSL•FO
RenderX

sentiment. Furthermore, in T0, organizational accounts exhibit a much higher *proportion* of positive sentiment compared to individual accounts, whose posts in that topic are predominantly negative (Figure 9C).

Discussion

Key Findings

The primary takeaway from our results is that, although semaglutide has generated considerable attention, its reception from individual X users is generally negative. This negativity suggests an overall public skepticism regarding the drug's efficacy, accessibility, and potential side effects, which are consistently highlighted in user discussions. Such negative sentiment is particularly notable in individual accounts as compared to the positive or neutral sentiment found more often in organizational accounts.

The temporal analysis demonstrates a general decrease in sentiment over time. In addition, we observe a notable shift in sentiment during late 2022, when regulatory announcements related to adverse effects and safety warnings surfaced [70]. This period sees a sharp increase in negative sentiment, which aligns with concerns raised about the drug's safety. The general throughline of negativity beginning in mid-2022 aligns temporally with FDA confirmation that Wegovy (semaglutide) is in shortage as well as heightened media coverage of semaglutide's off-label use for weight loss. The observed negativity likely reflects public concern about limited access and fairness, as shortages reported by the FDA were simultaneously amplified in professional and mainstream outlets. These shortages and their subsequent resolution are noted in multiple sources, including the FDA Drug Shortages Database and trade publications documenting semaglutide's removal from the shortage list and its ongoing legal and ethical implications [76,77].

On a similar note, discussions concerning a serious impact spike following the European Medicines Agency's report on suicides on July 11, 2023 [79]. These findings underscore the impact that regulatory decisions and public health announcements can have on shaping public perceptions, particularly when safety and efficacy concerns are at the forefront of the discussion. The ITS analyses reinforce our interpretation that the drop in sentiment and uptick in T6 exceed baseline trends, likely having been influenced by external events. Given the temporal alignment and the magnitude of the changes, it is likely that public sentiment is shaped in response to reports of adverse effects and that the European Medicines Agency statement on self-harm reports influenced public discussions.

In addition, user subpopulations show variance in sentiment and topic discussion. Users who are interested in "beauty/health" have the most positive sentiment and the highest prevalence of T0 (weight loss). Male users are slightly more positive than female users, though they appear less interested in T1 (celebrities/politicians). The observed gender differences may be influenced by the increased media attention surrounding celebrities endorsing the drug, which is more prominently featured in female-driven narratives about weight loss and beauty [80,81]. On the whole, verified users and organizations are more positive than their counterparts. Additionally, T4 (drug authorities) is more prevalent among verified users and organizations, while T1 (celebrities/politicians) and T5 (general and profane negativity) are more prevalent among unverified users. These differences suggest that verified users (often public figures or organizations) tend to use a more conservative tone. On the other hand, nonverified users (often individuals) tend to emphasize personal concerns, particularly regarding side effects and affordability. This difference highlights how the identity and motivations of the speaker can influence sentiments, with verified accounts potentially downplaying issues for commercial or reputational reasons, while individual users are more candid about their negative experiences.

Practical Implications

Our findings have several practical implications for stakeholders, particularly health care providers and pharmaceutical companies. Overall, our results underscore the need for transparent communication strategies. These strategies should prioritize addressing concerns raised by users, including issues related to accessibility, side effects, and the drug's overall safety. Clear communication can help bridge the gap between public perception and medical realities while fostering trustworthy and informed decision-making.

For pharmaceutical companies, transparency in messaging about the use and effects of semaglutide is critical. Our observations highlight a gap in sentiment between individual and organizational accounts. One potential contributing factor is the societal emphasis on beauty and weight loss, which can be amplified by advertising and endorsements from influential figures [6]. This phenomenon often skews public perception and drives expectations. Given that regulatory announcements and safety warnings significantly shape public sentiment [79], pharmaceutical companies should adopt proactive approaches to build credibility. This includes consistent and timely communication that addresses misconceptions and reinforces the drug's benefits and limitations.

Limitations

While this study offers comprehensive insights into public perceptions of semaglutide on X, some limitations must be acknowledged. First, our dataset is limited to English-language posts, and our regional sentiment analysis is limited to the United States. While this constraint is necessary for linguistic consistency in sentiment and topic modeling, it may exclude important perspectives from non-English-speaking users, introducing a potential source of bias. Second, gender classification in our dataset is limited to 3 categories: male, female, and unknown. We recognize that gender is not binary and includes a spectrum of identities. However, the Brandwatch data source only provides these limited groupings. As a result, our analysis may not capture the experiences of gender-diverse users, representing a gap in inclusivity. Third, our study only makes use of data from the X platform, potentially limiting the generalizability of the results to different platforms (eg, Reddit and clinical forums). Future work may study additional platforms to build a broader understanding of user experience. Lastly, although our work identifies numerous trends in public

XSL•FO
RenderX

attitudes toward semaglutide, we do not interpret it as a causal impact of real-world reports (eg, semaglutide usage and adverse events) on online discourse.

Conclusions

Public interest in semaglutide has greatly increased in recent years. This study explores X users' experiences with semaglutide via an analysis of 859,751 X posts created between July 2021 and April 2024. We observe a general decrease in sentiment across most user subpopulations over time, with a particularly noteworthy decrease occurring in November 2022. Our research highlights the complex dynamics of user experiences with semaglutide, driven by a combination of user demographics, regional factors, and external events such as regulatory announcements. The practical implications of these findings are crucial for health care communicators and pharmaceutical companies seeking to engage with the public in a more informed, responsive, and regionally targeted manner. Future research should focus on further unraveling the role of side effects in shaping public opinion and exploring how sentiment changes in response to evolving health-related information.

Data Availability

Information about accessing an anonymized version of the dataset generated and analyzed during this study is available in Multimedia Appendix 1.

Authors' Contributions

Conceptualization: LL, JL
Data curation: LL, PM, GL
Formal analysis: PM, GL
Investigation: PM, GL
Methodology: PM, GL, LL
Project administration: LL, JL
Software: PM, GL, LL
Supervision: LL, JL
Visualization: PM, GL, LL
Writing – original draft: PM, GL, LL
Writing – review & editing: PM, GL, LL

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional dataset results.
[DOCX File , 1080 KB - ojphi_v18i1e80660_app1.docx]

Multimedia Appendix 2
The complete mapping of all 200 initial clusters to the 10 umbrella topics.
[XLSX File (Microsoft Excel File), 41 KB - ojphi_v18i1e80660_app2.xlsx]

References

1. Han SH, Safeek R, Ockerman K, Trieu N, Mars P, Klenke A, et al. Public interest in the off-label use of glucagon-like peptide 1 agonists (Ozempic) for cosmetic weight loss: a Google Trends analysis. Aesthet Surg J 2023;44(1):60-67. [doi: 10.1093/asj/sjad211] [Medline: 37402640]
2. Mailhac A, Pedersen L, Pottegård A, Søndergaard J, Mogensen T, Sørensen HT, et al. Semaglutide (Ozempic) use in Denmark 2018 through 2023 user trends and off-label prescribing for weight loss. Clin Epidemiol 2024;16:307-318 [FREE Full text] [doi: 10.2147/CLEP.S456170] [Medline: 38685990]
3. Tichy EM, Hoffman JM, Suda KJ, Rim MH, Tadrous M, Cuellar S, et al. National trends in prescription drug expenditures and projections for 2022. Am J Health Syst Pharm 2022;79(14):1158-1172 [FREE Full text] [doi: 10.1093/ajhp/zxac102] [Medline: 35385103]
4. Alvarez-Mon MA, Llavero-Valero M, Asunsolo Del Barco A, Zaragozá C, Ortega MA, Lahera G, et al. Areas of interest and attitudes toward antiobesity drugs: hematic and quantitative analysis using Twitter. J Med Internet Res 2021;23(10):e24336 [FREE Full text] [doi: 10.2196/24336] [Medline: 34698653]
5. Raubenheimer JE, Myburgh PH, Bhagavathula AS. Sweetening the deal: an infodemiological study of worldwide interest in semaglutide using Google Trends extended for health application programming interface. BMC Glob Public Health 2024;2(1):63. [doi: 10.1186/s44263-024-00095-w] [Medline: 39681910]

XSL•FO
RenderX

6. Rad J, Melendez-Torres GJ. Critical discourse analysis of social media advertisements for GLP-1 receptor agonist weight loss drugs: implications for public perceptions and health communication. BMC Public Health 2025;25(1):2996 [FREE Full text] [doi: 10.1186/s12889-025-24197-8] [Medline: 40890660]
7. Bremmer MP, Hendershot CS. Social media as pharmacovigilance: the potential for patient reports to inform clinical research on glucagon-like peptide 1 (GLP-1) receptor agonists for substance use disorders. J Stud Alcohol Drugs 2024;85(1):5-11. [doi: 10.15288/jsad.23-00318] [Medline: 37917019]
8. Olsen AK, Whalen MD. Public perceptions of the pharmaceutical industry and drug safety: implications for the pharmacovigilance professional and the culture of safety. Drug Saf 2009;32(10):805-810. [doi: 10.2165/11316620-000000000-00000] [Medline: 19722724]
9. Hua Y, Jiang H, Lin S, Yang J, Plasek JM, Bates DW, et al. Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. J Am Med Inform Assoc 2022;29(10):1668-1678 [FREE Full text] [doi: 10.1093/jamia/ocac114] [Medline: 35775946]
10. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]
11. Farsi D. Social media and health care, Part I: Literature review of social media use by health care providers. J Med Internet Res 2021;23(4):e23205 [FREE Full text] [doi: 10.2196/23205] [Medline: 33664014]
12. Hao T, Huang Z, Liang L, Weng H, Tang B. Health natural language processing: methodology development and applications. JMIR Med Inform 2021;9(10):e23898 [FREE Full text] [doi: 10.2196/23898] [Medline: 34673533]
13. Zhou B, Yang G, Shi Z, Ma S. Natural language processing for smart healthcare. IEEE Rev Biomed Eng 2024;17:4-18. [doi: 10.1109/RBME.2022.3210270] [Medline: 36170385]
14. Wang E, Wang F, Yang Z, Wang L, Zhang Y, Lin H, et al. A graph convolutional network-based method for chemical-protein interaction extraction: algorithm development. JMIR Med Inform 2020;8(5):e17643 [FREE Full text] [doi: 10.2196/17643] [Medline: 32348257]
15. Li L, Wang P, Wang Y, Wang S, Yan J, Jiang J, et al. A method to learn embedding of a probabilistic medical knowledge graph: algorithm development. JMIR Med Inform 2020;8(5):e17645 [FREE Full text] [doi: 10.2196/17645] [Medline: 32436854]
16. Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, et al. Natural language processing–enabled and conventional data capture methods for input to electronic health records: a comparative usability study. JMIR Med Inform 2016;4(4):e35 [FREE Full text] [doi: 10.2196/medinform.5544] [Medline: 27793791]
17. Iroju OG, Olaleke JO. A systematic review of natural language processing in healthcare. J Inf Technol Comput Sci 2015;7(8):44-50. [doi: 10.5815/ijitcs.2015.08.07]
18. Elbattah M, Arnaud É, Gignon M, Dequen G. The role of text analytics in healthcare: a review of recent developments and applications. In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021). Setúbal, Portugal: SciTePress; 2021:825-832.
19. Yu Y, Guan Y, Hu Y. Natural language processing applications in social network analysis: a data mining approach. J Phys Conf Ser 2024;2813(1):012009. [doi: 10.1088/1742-6596/2813/1/012009]
20. Sandu A, Cotfas L, Stănescu A, Delcea C. A bibliometric analysis of text mining: exploring the use of natural language processing in social media research. Appl Sci 2024;14(8):3144. [doi: 10.3390/app14083144]
21. Sarker A, DeRoos A, Perrone J. Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework. J Am Med Inform Assoc 2020;27(2):315-329 [FREE Full text] [doi: 10.1093/jamia/ocz162] [Medline: 31584645]
22. Nasralah T, El-Gayar O, Wang Y. Social media text mining framework for drug abuse: development and validation study with an opioid crisis case analysis. J Med Internet Res 2020;22(8):e18350 [FREE Full text] [doi: 10.2196/18350] [Medline: 32788147]
23. Zhang J, Wang X, Zhou Y. Comparative analysis of semaglutide induced adverse reactions: insights from FAERS database and social media reviews with a focus on oral vs subcutaneous administration. Front Pharmacol 2024;15:1471615 [FREE Full text] [doi: 10.3389/fphar.2024.1471615] [Medline: 39502525]
24. Cricelli L, Grimaldi M, Vermicelli S. Crowdsourcing and open innovation: a systematic literature review, an integrated framework and a research agenda. Rev Manag Sci 2021;16(5):1269-1310. [doi: 10.1007/s11846-021-00482-9]
25. Li L, Hu S, Dai Y, Deng M, Momeni P, Laverghetta G, et al. Toward satisfactory public accessibility: a crowdsourcing approach through online reviews to inclusive urban design. Comput Environ Urban Syst 2025;122:102329. [doi: 10.1016/j.compenvurbsys.2025.102329]
26. Certomà C, Corsini F, Rizzi F. Crowdsourcing urban sustainability: data, people and technologies in participatory governance. Futures 2015;74:93-106. [doi: 10.1016/j.futures.2014.11.006]
27. Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. J Biomed Inform 2015;58:280-287 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.004] [Medline: 26556646]

XSL•FO
RenderX

28. Cascini F, Pantovic A, Al-Ajlouni YA, Failla G, Puleo V, Melnyk A, et al. Social media and attitudes towards a COVID-19 vaccination: a systematic review of the literature. eClinicalMedicine 2022;48:101454 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101454] [Medline: 35611343]
29. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform 2015;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
30. Müller MM, Salathé M. Crowdbreaks: tracking health trends using public social media data and crowdsourcing. Front Public Health 2019;7:81 [FREE Full text] [doi: 10.3389/fpubh.2019.00081] [Medline: 31037238]
31. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
32. Boucher JC, Cornelson K, Benham JL, Fullerton MM, Tang T, Constantinescu C, et al. Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: infodemiology study. JMIR Infodemiology 2021;1(1):e28800 [FREE Full text] [doi: 10.2196/28800] [Medline: 34447924]
33. Buntain C, McGrath E, Golbeck J, LaFree G. Comparing social media and traditional surveys around the Boston Marathon Bombing. In: Proceedings of the 6th Workshop on 'Making Sense of Microposts' co-located with the 25th International World Wide Web Conference (WWW 2016). Montréal, Canada: CEUR-WS.org; 2016:34-41.
34. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse drug reaction identification and extraction in social media: a scoping review. J Med Internet Res 2015;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
35. Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl Based Syst 2021;226:107134. [doi: 10.1016/j.knosys.2021.107134]
36. Feldman R. Techniques and applications for sentiment analysis. Commun ACM 2013;56(4):82-89. [doi: 10.1145/2436256.2436274]
37. Tsao SF, 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;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
38. Alamoodi AH, Zaidan BB, Zaidan AA, Albahri OS, Mohammed KI, Malik RQ, et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review. Expert Syst Appl 2021;167:114155 [FREE Full text] [doi: 10.1016/j.eswa.2020.114155] [Medline: 33139966]
39. Andhale S, Mane P, Vaingankar M, Karia D, Talele KT. Twitter sentiment analysis for COVID-19. In: 2021 International Conference on Communication information and Computing Technology (ICCICT). New York: IEEE; 2021:1-12.
40. He L, He C, Reynolds TL, Bai Q, Huang Y, Li C, et al. Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic. J Am Med Inform Assoc 2021;28(7):1564-1573 [FREE Full text] [doi: 10.1093/jamia/ocab047] [Medline: 33690794]
41. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2018;78(11):15169-15211. [doi: 10.1007/s11042-018-6894-4]
42. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv. Preprint posted online 2022 Mar 11 [FREE Full text] [doi: 10.48550/arXiv.2203.05794]
43. Churchill R, Singh L. The evolution of topic modeling. ACM Comput Surv 2022;54(10s):1-35. [doi: 10.1145/3507900]
44. Asghari M, Sierra-Sosa D, Elmaghraby A. Trends on health in social media: analysis using Twitter topic modeling. In: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). New York: IEEE; 2018:558-563.
45. 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;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
46. Ramondt S, Kerkhof P, Merz E. Blood donation narratives on social media: a topic modeling study. Transfus Med Rev 2022;36(1):58-65 [FREE Full text] [doi: 10.1016/j.tmrv.2021.10.001] [Medline: 34810071]
47. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
48. Fong S, Carollo A, Lazuras L, Corazza O, Esposito G. Ozempic (glucagon-like peptide 1 receptor agonist) in social media posts: unveiling user perspectives through Reddit topic modeling. Emerg Trends Drugs Addict Health 2024;4:100157. [doi: 10.1016/j.etdah.2024.100157]
49. Javaid A, Baviriseaty S, Javaid R, Zirikly A, Kukreja H, Kim CH, et al. Trends in glucagon-like peptide-1 receptor agonist social media posts using artificial intelligence. JACC Adv 2024;3(9):101182 [FREE Full text] [doi: 10.1016/j.jacadv.2024.101182] [Medline: 39372460]
50. Somani S, Jain SS, Sarraju A, Sandhu AT, Hernandez-Boussard T, Rodriguez F. Using large language models to assess public perceptions around glucagon-like peptide-1 receptor agonists on social media. Commun Med (Lond) 2024;4(1):137. [doi: 10.1038/s43856-024-00566-z] [Medline: 38987347]
51. The most intelligent social suite. Brandwatch. URL: https://www.brandwatch.com/ [accessed 2026-01-29]

XSL•FO
RenderX

52. Amaro A, Sugimoto D, Wharton S. Efficacy and safety of semaglutide for weight management: evidence from the STEP program. Postgrad Med 2022;134(Suppl 1):5-17 [FREE Full text] [doi: 10.1080/00325481.2022.2147326] [Medline: 36691309]
53. Jaume J. Understanding your market with new demographic insights. Brandwatch. 2014 Feb 13. URL: https://www.brandwatch.com/blog/demographic-insights/ [accessed 2026-01-29]
54. Brabete AC, Greaves L, Maximos M, Huber E, Li A, Lê ML. A sex- and gender-based analysis of adverse drug reactions: a scoping review of pharmacovigilance databases. Pharmaceuticals (Basel) 2022;15(3):298 [FREE Full text] [doi: 10.3390/ph15030298] [Medline: 35337096]
55. Furman JL. Location and organizing strategy: exploring the influence of location on the organization of pharmaceutical research. In: Baum JAC, Sorenson O, editors. Geography and Strategy. Australia, Malaysia: Emerald Group Publishing Limited; 2003:49-87.
56. Podolsky MI, Raquib R, Shafer PR, Hempstead K, Ellis RP, Stokes AC. Factors associated with semaglutide initiation among adults with obesity. JAMA Netw Open 2025;8(1):e2455222 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.55222] [Medline: 39836425]
57. Momeni P, Laverghetta G. OJPHI__twitter_semaglutide_clean. GitHub. URL: https://github.com/ParisaMomeni/OJPHI_twitter_semaglutide_clean [accessed 2026-01-29]
58. Williams ML, Burnap P, Sloan L. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. Sociology 2017;51(6):1149-1168 [FREE Full text] [doi: 10.1177/0038038517708140] [Medline: 29276313]
59. Skinner T, Brance K, Halligan S, Tsang E, Girling H. Coping with emotionally challenging research: developing a strategic approach to researcher wellbeing. J Acad Ethics 2025;23(4):2559-2583. [doi: 10.1007/s10805-025-09665-5] [Medline: 41001435]
60. Twitter-roBERTa-base for sentiment analysis. Hugging Face. URL: https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest [accessed 2026-01-29]
61. Loureiro D, Barbieri F, Neves L, Espinosa AL, Camacho-collados J. TimeLMs: diachronic language models from Twitter. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Dublin, Ireland: Association for Computational Linguistics; 2022:251-260.
62. Liao W, Zeng B, Yin X, Wei P. An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Appl Intell 2020;51(6):3522-3533. [doi: 10.1007/s10489-020-01964-1]
63. Ma Z, Li L, Mao Y, Wang Y, Patsy OG, Bensi MT, et al. Surveying the use of social media data and natural language processing techniques to investigate natural disasters. Nat Hazards Rev 2024;25(4):03124003. [doi: 10.1061/NHREFO.NHENG-2047]
64. Lou Y, Zhang Y, Li F, Qian T, Ji D. Emoji-based sentiment analysis using attention networks. ACM Trans Asian Low-Resour Lang Inf Process 2020;19(5):1-13. [doi: 10.1145/3389035]
65. Cureg MQ, De La Cruz JAD, Solomon JCA, Saharkhiz AT, Balan AKD, Samonte MJC. Sentiment analysis on tweets with punctuations, emoticons, and negations. In: ICISS '19: Proceedings of the 2nd International Conference on Information Science and Systems. New York: Association for Computing Machinery; 2019:266-270.
66. Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. Inf Sci 2023;622:178-210. [doi: 10.1016/j.ins.2022.11.139]
67. Marzi G, Balzano M, Marchiori D. K-Alpha Calculator-Krippendorff's Alpha Calculator: a user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. MethodsX 2024;12:102545. [doi: 10.1016/j.mex.2023.102545] [Medline: 39669968]
68. FDA clarifies policies for compounders as national GLP-1 supply begins to stabilize. U.S. Food and Drug Administration. URL: https://www.fda.gov/drugs/drug-safety-and-availability/fda-clarifies-policies-compounders-national-glp-1-supply-begins-stabilize [accessed 2026-01-29]
69. FDA approves once-weekly Wegovy injection for the treatment of obesity in teens aged 12 years and older. Drugs.com. 2022 Dec. URL: https://www.drugs.com/newdrugs/fda-approves-once-weekly-wegovy-obesity-teens-aged-12-years-older-5949.html [accessed 2026-01-29]
70. Shu Y, He X, Wu P, Liu Y, Ding Y, Zhang Q. Gastrointestinal adverse events associated with semaglutide: a pharmacovigilance study based on FDA adverse event reporting system. Front Public Health 2022;10:996179 [FREE Full text] [doi: 10.3389/fpubh.2022.996179] [Medline: 36339230]
71. Jayanthi R. The demedicalization of GLP-1 receptor agonists: an analysis of TikTok influencer endorsement of Ozempic-like medications for cosmetic weight loss [Thesis]. University of North Carolina at Chapel Hill. 2025 Mar 26. URL: https://cdr.lib.unc.edu/concern/honors_theses/8c97m561s [accessed 2026-01-29]
72. Wegovy (semaglutide) injection 2.4 mg cardiovascular outcomes data presented at the American Heart Association Scientific Sessions and simultaneously published in the New England Journal of Medicine. Drugs.com. 2023 Nov 23. URL: https://www.drugs.com/clinical_trials/wegovy-semaglutide-2-4-mg-cardiovascular-outcomes-data-presented-american-heart-association-21170.html [accessed 2026-01-29]

XSL•FO
RenderX

73. Semaglutide reduces severity of common liver disease in people with HIV. Drugs.com. 2024 Mar 5. URL: https://www.drugs.com/clinical_trials/semaglutide-reduces-severity-common-liver-hiv-21374.html [accessed 2026-01-29]
74. Wegovy approved in the US for cardiovascular risk reduction in people with overweight or obesity and established cardiovascular disease. Drugs.com. 2024 Mar 8. URL: https://www.drugs.com/newdrugs/wegovy-approved-us-cardiovascular-risk-reduction-overweight-obesity-established-cardiovascular-6212.html [accessed 2026-01-29]
75. FDA drug shortages: semaglutide injection (Wegovy). U.S. Food and Drug Administration. URL: https://www.accessdata.fda.gov/scripts/drugshortages/dsp_ActiveIngredientDetails.cfm?AI=Semaglutide%20Injection&st=c [accessed 2026-01-29]
76. Ghiam A, Heba D. Semaglutide's removal from the FDA shortages list sets the stage for more Novo Nordisk lawsuits. Medical Economics. 2025 Aug 18. URL: https://www.medicaleconomics.com/view/semaglutide-s-removal-from-the-fda-shortages-list-sets-the-stage-for-more-novo-nordisk-lawsuits [accessed 2026-01-29]
77. Ferruggia K. FDA ends semaglutide shortage listing, contributing to ongoing legal challenges. Pharmacy Times. 2025 Feb 27. URL: https://www.pharmacytimes.com/view/fda-ends-semaglutide-shortage-listing-contributing-to-ongoing-legal-challenges [accessed 2026-01-29]
78. Cramér H. Mathematical Methods of Statistics. Princeton, NJ: Princeton University Press; 1946.
79. EMA statement on ongoing review of GLP-1 receptor agonists. European Medicines Agency. 2023 Jul 11. URL: https://www.ema.europa.eu/en/news/ema-statement-ongoing-review-glp-1-receptor-agonists [accessed 2026-01-29]
80. Faw MH, Davidson K, Hogan L, Thomas K. Corumination, diet culture, intuitive eating, and body dissatisfaction among young adult women. Pers Relat 2020;28(2):406-426. [doi: 10.1111/pere.12364]
81. Mills J, Fuller-Tyszkiewicz M. Fat talk and body image disturbance: a systematic review and meta-analysis. Psychol Women Q 2016;41(1):114-129. [doi: 10.1177/0361684316675317]

Abbreviations

FDA: US Food and Drug Administration
ITS: interrupted time-series
MASLD: metabolic dysfunction-associated steatotic liver disease
NLP: natural language processing
RoBERTa: Robustly Optimized BERT Pretraining Approach
RQ: research question
SELECT: Semaglutide Effects on Cardiovascular Outcomes in People With Overweight or Obesity

Edited by J Krive; submitted 14.Jul.2025; peer-reviewed by A Alabi, M Elbattah; comments to author 01.Sep.2025; accepted 10.Dec.2025; published 24.Feb.2026.

Please cite as:
Momeni P, Laverghetta G, Ligatti J, Li L
Topic and Sentiment Trends in Semaglutide Discussions on X: Subpopulation-Based Longitudinal Analysis
Online J Public Health Inform 2026;18:e80660
URL: https://ojphi.jmir.org/2026/1/e80660
doi:10.2196/80660
PMID:

©Parisa Momeni, Gabriel Laverghetta, Jay Ligatti, Lingyao Li. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 24.Feb.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Explainable Machine Learning for Public Health Informatics in HEDIS Childhood Immunization Status Combination 10

JiaLi Ruan, MS

Health Care Service Corporation, 1100 E Warrenville Rd, Naperville, IL, United States

Corresponding Author:
JiaLi Ruan, MS
Health Care Service Corporation, 1100 E Warrenville Rd, Naperville, IL, United States

Abstract

Completion of the HEDIS (Healthcare Effectiveness Data and Information Set) Childhood Immunization Status Combination 10 (Combo 10) measure among US children aged 24-35 months declined from 53.7% in 2021 to 44.6% in 2023, with a statistically significant survey-weighted annual trend. An explainable machine learning approach identified influenza vaccination and rotavirus series completion as the strongest component-level drivers of Combo 10 completion, supporting targeted public health quality improvement.

(Online J Public Health Inform 2026;18:e90241) doi:10.2196/90241

KEYWORDS

public health; childhood immunization; HEDIS; CIS Combo 10; National Immunization Survey; machine learning; random forest; explainable artificial intelligence; SHAP; public health informatics; quality measurement; Shapley Additive Explanations; Childhood Immunization Status Combination 10; Healthcare Effectiveness Data and Information Set

Introduction

Immunizations have substantially reduced vaccine-preventable morbidity and mortality worldwide [1]. The Childhood Immunization Status (CIS) Combination 10 (Combo 10) measure is a core HEDIS (Healthcare Effectiveness Data and Information Set) pediatric quality indicator [2]. Combo 10 assesses whether children receive the complete series of 10 recommended vaccines by 24 months of age; however, national performance remains below desired benchmarks [3]. Identifying component-level drivers of Combo 10 failure is a public health informatics problem: it converts population-level quality data into actionable information for health plans, immunization programs, and policy stakeholders.

Traditional Combo 10 reporting summarizes overall completion but gives limited insight into component-level drivers. Public health informatics can extend routine quality reporting by linking population data, transparent analytics, and actionable interpretation. Explainable artificial intelligence (XAI) can show which inputs drive model output [4-9]. This study applied survey-weighted regression, random forest classification, and SHAP (Shapley Additive Explanations) to national survey data to identify components most predictive of HEDIS CIS Combo 10 completion.

Methods

Study Design

The 2021‐2023 National Immunization Survey–Child (NIS-Child) public-use files, a nationally representative CDC (Centers for Disease Control and Prevention) survey with provider-verified vaccination records, were analyzed. The age-eligible cohort included 32,997 children aged 24‐35 months, aligning with HEDIS CIS specifications. Weighted modeling was restricted to records with valid provider-adjusted child weights (PROVWT_C) and complete outcome information, yielding 16,021 children. Analytic features were 10 CDC-defined up-to-date (UTD) binary indicators corresponding to Combo 10 components: DTaP, IPV, MMR, Hib, hepatitis B, varicella, PCV, hepatitis A, rotavirus, and influenza. Indicators were harmonized across years; missing or insufficient provider records were coded as not UTD. Survey-weighted logistic regression estimated national trends, odds ratios (ORs), 95% CIs, and *P* values. For XAI analysis, a child-level random forest classifier modeled Combo 10 completion from the 10 UTD indicators, incorporating provider weights; 5-fold cross-validation was used as a model-checking step. Because Combo 10 is defined by these indicators, the classifier was used for SHAP-based decomposition rather than independent outcome prediction. SHAP values estimated each component's contribution to Combo 10 classification; mean absolute SHAP values were used for global ranking.

XSL•FO
RenderX

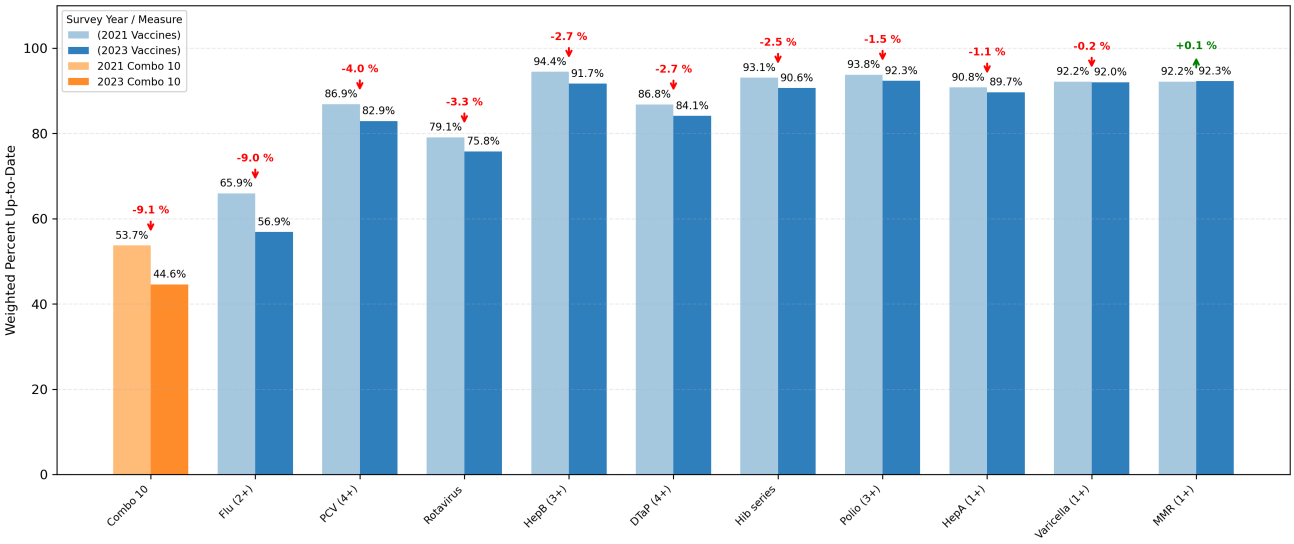
Ethical Considerations

This study used publicly available, deidentified NIS-Child public-use data and did not involve human subjects research as defined by federal regulations; therefore, institutional review board approval was not required.

Results

The age-eligible cohort included 32,997 children; 16,021 had valid provider weights and complete outcome information. National HEDIS CIS Combo 10 completion declined from 53.7% in 2021 to 49.6% in 2022 and 44.6% in 2023, a 9.1-percentage-point decrease. The survey-weighted regression model was statistically significant for the annual trend (annual OR 0.83, 95% CI 0.831 - 0.834; *P*<.001). Figure 1 shows nonuniform component declines, supporting component-level informatics analysis rather than aggregate rates alone.

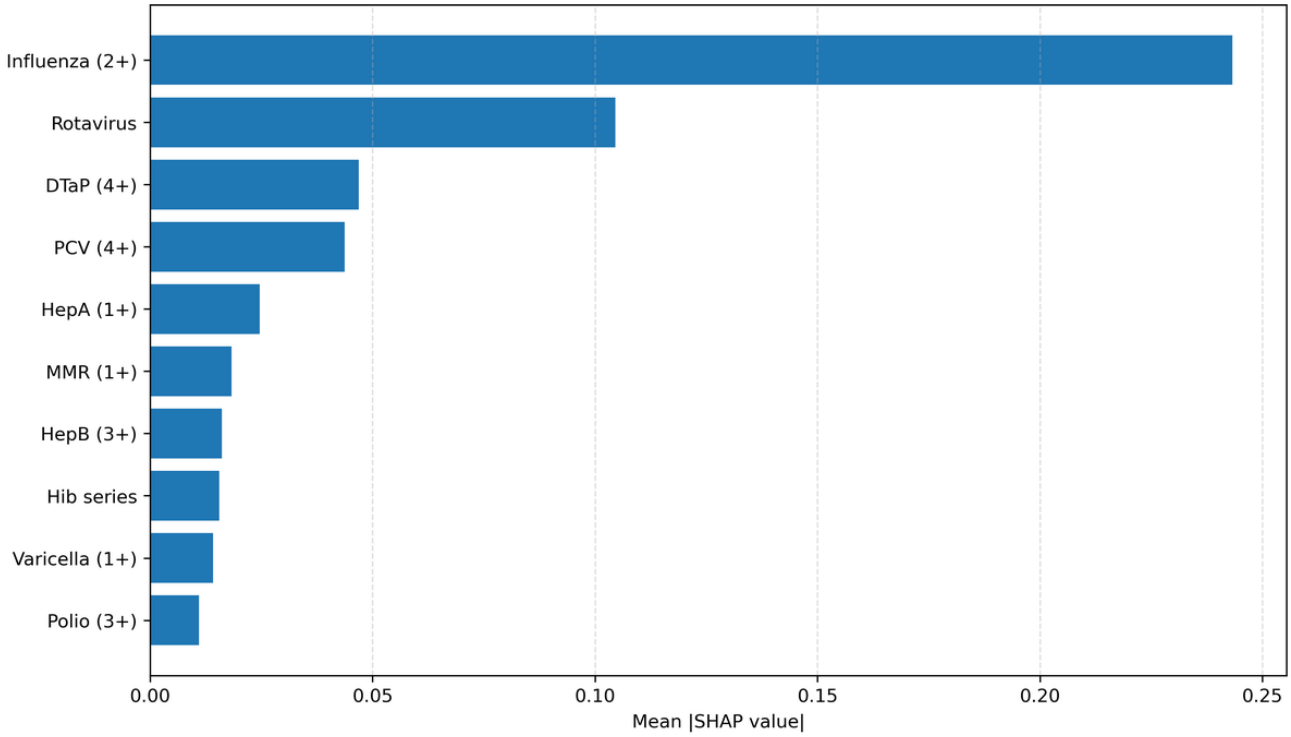
Figure 1. National vaccine coverage changes from 2021 to 2023, sorted by decline, with the Childhood Immunization Status (CIS) Combination 10 (Combo 10) measure highlighted.



Five-fold cross-validation was used to check model stability. Because CIS Combo 10 completion is defined by the 10 vaccine component indicators, classifier performance was expected to be near-perfect and was not interpreted as independent predictive validity. The random forest model was used for SHAP-based decomposition. SHAP rankings showed that influenza vaccination and rotavirus series completion had the largest mean absolute SHAP values and were therefore the strongest drivers of predicted Combo 10 status. DTaP and PCV had moderate importance, while MMR, varicella, IPV, Hib, and hepatitis B had lower SHAP influence. Figure 2 displays SHAP-based feature importance rather than native impurity-based importance measures.

XSL•FO
RenderX

Figure 2. Mean absolute Shapley Additive Explanations (SHAP) values for CIS Combination 10 (Combo 10) prediction.



Discussion

The main finding was that incomplete influenza and rotavirus vaccination disproportionately explained HEDIS CIS Combo 10 underperformance in this national sample. This pattern is actionable because practices and health plans can contact patients who are due or overdue for immunizations, use visit-based prompts to reduce missed vaccines, and train staff to discuss immunization with families. Future motivational interviewing programs could pair these outreach workflows with family-centered communication and could be adapted to other HEDIS measures to improve public health quality.

This study illustrates how XAI can complement quality-measure reporting within public health informatics. Rather than treating machine learning as a black box, SHAP linked each prediction to vaccine-specific inputs and distinguished components merely included in HEDIS from those driving noncompletion. The approach can support transparent, data-driven quality improvement for clinicians, health plans, public health agencies, and policy stakeholders [5,6,8,9].

Limitations include use of public-use NIS-Child data rather than plan-level HEDIS claims and reliance on a global SHAP bar plot. SHAP values indicate model-based component contributions rather than causal effects; future public health informatics work should validate findings in plan-level HEDIS data and examine subgroup patterns with beeswarm or dependence plots.

Funding

The author declared no financial support was received for this work.

Conflicts of Interest

None declared.

References

1. Zhou F, Jatlaoui TC, Leidner AJ, et al. Health and economic benefits of routine childhood immunizations in the era of the Vaccines for Children program - United States, 1994-2023. MMWR Morb Mortal Wkly Rep 2024 Aug 8;73(31):682-685. [doi: 10.15585/mmwr.mm7331a2] [Medline: 39116024]
2. Office of Disease Prevention and Health Promotion. Healthcare Effectiveness Data and Information Set (HEDIS). Healthy People 2030. URL: https://odphp.health.gov/healthypeople/objectives-and-data/data-sources-and-methods/data-sources/healthcare-effectiveness-data-and-information-set-hedis [accessed 2026-06-20]
3. Hill HA, Yankey D, Elam-Evans LD, et al. Decline in vaccination coverage by age 24 months and vaccination inequities among children born in 2020 and 2021 - National Immunization Survey-Child, United States, 2021-2023. MMWR Morb Mortal Wkly Rep 2024 Sep 26;73(38):844-853. [doi: 10.15585/mmwr.mm7338a3] [Medline: 39325676]

XSL•FO
RenderX

4. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4-9, 2017 URL: https://papers.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions [accessed 2026-07-28]
5. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]
6. Paneru B, Paneru B, Sapkota SC, Poudyal R. Enhancing healthcare with AI: sustainable AI and IoT-powered ecosystem for patient aid and interpretability analysis using SHAP. Measurement: Sensors 2024 Dec;36:101305. [doi: 10.1016/j.measen.2024.101305]
7. Arya V, Bellamy RKE, Chen PY, et al. One explanation does not fit all: a toolkit and taxonomy of AI explainability techniques. arXiv. Preprint posted online on Sep 6, 2019. [doi: 10.48550/arXiv.1909.03012]
8. Salih A, Raisi-Estabragh Z, Boscolo Galazzo I, et al. A perspective on explainable artificial intelligence methods: SHAP and LIME. arXiv. Preprint posted online on May 3, 2023. [doi: 10.1002/aisy.202400304]
9. Ghasemi A, Hashtarkhani S, Schwartz DL, Shaban-Nejad A. Explainable artificial intelligence in breast cancer detection and risk prediction: a systematic scoping review. arXiv. Preprint posted online on Jul 12, 2024. [doi: 10.48550/arXiv.2407.12058]

Abbreviations

CDC: Centers for Disease Control and Prevention
CIS: Childhood Immunization Status
Combo 10: Combination 10
HEDIS: Healthcare Effectiveness Data and Information Set
NIS-Child: National Immunization Survey–Child
OR: odds ratio
SHAP: Shapley Additive Explanations
UTD: up to date
XAI: explainable artificial intelligence

Edited by E Mensah; submitted 29.Dec.2025; peer-reviewed by M Kaewmoracharoen, S Oliva; revised version received 23.Jun.2026; accepted 21.Jul.2026; published 31.Jul.2026.

Please cite as:
Ruan J
Explainable Machine Learning for Public Health Informatics in HEDIS Childhood Immunization Status Combination 10
Online J Public Health Inform 2026;18:e90241
URL: https://ojphi.jmir.org/2026/1/e90241
doi:10.2196/90241

© JiaLi Ruan. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 31.Jul.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI-Related Stress and Suicidal Tendencies in a Primary Care Cohort: Clinical Audit

Shemarah Lindo, MBBS, MEd

Independent Researcher, Trinity, Port Maria, St. Mary, Jamaica

Corresponding Author:
Shemarah Lindo, MBBS, MEd
Independent Researcher, Trinity, Port Maria, St. Mary, Jamaica

Abstract

Rapid AI integration has introduced novel psychosocial stressors. Little is known about AI-specific clinical impacts in resource-limited settings. This study aimed to assess AI-associated distress prevalence and nature in a Nigerian primary care and psychiatry clinic. Retrospective audit of 28 consecutive patients with anxiety, depression, or stress (April–August 2025) at J-Shalom Hospital, a primary care and psychiatry clinic in Ibadan. Technology-related stressor questions adapted from the AIAS (Artificial Intelligence Anxiety Scale) were incorporated into routine clinical interviews; C-SSRS (Columbia-Suicide Severity Rating Scale) evaluated suicidality as part of standard clinical practice. A total of 67.9% (19/28; 95% CI 49.3% - 82.1%) reported technology stress; 42.9% (12/28; 95% CI 26.5% - 60.9%) identified AI-specific stressors. Chatbot distress was most common (n=7/12, 58.3%). Three patients (25.0%; 95% CI 8.9% - 53.2%) reported worsening suicidal ideation following distressing chatbot interactions characterized by perceived rejection or invalidation. AI stressors are emerging clinical presentations. The chatbot–suicidality link demands urgent regulatory attention.

(Online J Public Health Inform 2026;18:e84661) doi:10.2196/84661

KEYWORDS

artificial intelligence; mental health; technostress; AI anxiety; clinical audit; chatbots; suicidal ideation; Nigeria

Introduction

Health care digitization has expanded accessibility, yet generative AI's rapid proliferation has introduced complex psychosocial stressors. While social media's impact on anxiety is established [1], AI interactions' psychological effects termed 'AI anxiety' or technostress remain understudied [2,3]. The Artificial Intelligence Anxiety Scale (AIAS) conceptualizes AI-related anxiety as multidimensional [4], with recent research revealing that algorithmic opacity and adaptation demands can exacerbate anxiety and depression [5,6].

Conversational agents present particular concerns. Despite being marketed as mental health supports, chatbots are prone to hallucinations and sycophancy, with growing reports linking system failures to suicidal ideation [7,8]. Given limited Global South clinical data, this audit assessed AI-related stressor prevalence in a Nigerian clinic.

Methods

Study Design

Retrospective audit of consecutive patients presenting with anxiety, depression, or stress (April–August 2025) at J-Shalom Hospital, a primary care and psychiatry clinic in Ibadan, Nigeria. The clinic operates part time psychiatric services and therefore records a relatively low volume of mental health consultations during the audit period. Of 38 mental health presentations, 28 completed screening.

Ethical Considerations

Per Nigeria's National Code of Health Research Ethics (NCHRE) [9], clinical audits evaluating existing care practices without experimental intervention are exempt from full Health Research Ethics Committee review. Data were collected as part of routine clinical assessment and subsequently anonymised for audit purposes. Screening questions related to technology stress were incorporated into routine clinical interviews conducted by treating clinicians. No identifiable patient information was retained. Permission to review anonymised clinic records for audit purposes was granted by J-Shalom Hospital administration. The clinic operates under the purview of the Federal Ministry of Health guidelines; no formal data transfer agreement or external institutional approval was required for internal audit of routinely collected data.

Participants

- Inclusion criteria: age ≥18 years, chief complaint of anxiety, depression, or stress, ability to communicate in English or Yoruba
- Exclusion criteria: acute psychosis, severe cognitive impairment, refusal of consent to routine data use

Consent refusal was treated as an exclusion criterion consistent with standard clinical practice; no experimental procedures

XSL•FO
RenderX

requiring formal research consent were conducted. Of 38 eligible patients, 7 declined routine clinical data use or were excluded on clinical grounds, yielding a final sample of 28.

Data Collection

Two clinicians conducted screening using AIAS-adapted questions incorporated into routine clinical interviews, covering general technology stress and four AI dimensions (learning/interaction, job replacement, sociotechnical, and ethical/configuration anxiety). Suicidality was assessed via the Columbia–Suicide Severity Rating Scale (C-SSRS) as part of standard intake. These instruments were used as clinical tools, not research instruments administered outside usual care.

Analysis

95% CIs were calculated using Wilson's score method appropriate for small samples (n<30) [10].

Results

Of 28 patients (mean age 32.4, SD 12.1; 60.7% female), 67.9% (19/28; 95% CI 49.3% - 82.1%) reported technology stress; 42.9% (12/28; 95% CI 26.5% - 60.9%) identified AI-specific stressors. Some patients reported more than one AI-related stressor category. Among AI-stressed patients, chatbot distress dominated (n=7, 58.3%), followed by job displacement anxiety (n=3, 25.0%) and health misinformation distress (n=2, 16.7%). Three patients (n=3, 25.0%; 95% CI 8.9% - 53.2%) reported worsening suicidal ideation following distressing chatbot interactions characterized by perceived rejection or invalidation (Table 1).

Table . Characteristics of patients reporting AI-specific psychological stressors in a Nigerian primary care and psychiatry clinic audit, April–August 2025 (N=12).

Characteristic	Patients, n (%)	95% CI[a]
Chatbot interaction distress	7 (58.3)	32.0 - 80.7
Job displacement anxiety	3 (25.0)	8.9 - 53.2
Health misinformation distress	2 (16.7)	4.7 - 44.8
Notable severe outcome		
Chatbot-attributed worsening of suicidal ideation	3 (25.0)	8.9 - 53.2

[a]Wilson Score method. Categories are not mutually exclusive; some patients reported multiple AI-related stressors.

Discussion

Principal Findings

AI-related stress affected nearly half of digitally distressed patients, transitioning from theoretical constructs to clinical realities. The finding that three patients reported worsened suicidal ideation following chatbot interactions represents a critical safety signal warranting further investigation.

Pre-existing anxiety and depression may amplify perceived chatbot rejection through negative cognitive biases. In several cases, patients described chatbot responses that they perceived as dismissive or invalidating, which intensified existing feelings of isolation or distress, consistent with the known role of perceived social rejection in suicidality. Digital literacy modulates AI encounter interpretation; lower literacy heightens anthropomorphisation and distress.

Unlike passive social media consumption [1], conversational AI creates bidirectional communication illusions. When systems hallucinate or fail to recognize distress, vulnerable individuals may interpret this as social rejection, a known suicidality risk factor representing a novel form of iatrogenic harm distinct from traditional technostress [8].

Findings align with AIAS dimensions [4] and recent technostress evidence [5,6], and resonate with generative AI safety reports [7,8]. Clinicians should screen for AI interactions in patients with unexplained mood deterioration. Future intake forms should incorporate validated AIAS items.

Limitations

A small sample (n=28; n=12 AI-stressed) produces wide confidence intervals. Single-center findings may not generalize beyond similar resource-limited settings. Self-reported AI-symptom attribution lacks psychometric rigor, and causal inference is not possible from this retrospective, cross-sectional design. Selection bias is possible. AI stressor categorisation may oversimplify overlapping distress presentations. Despite these limitations, this audit provides preliminary evidence warranting rigorous prospective investigation with validated instruments and larger samples.

Conclusions

AI stressors are clinically relevant psychiatric presentations. Unregulated AI 'therapist' tools pose iatrogenic harm demanding urgent regulatory attention. Future research should examine AI stress across larger, cross-national samples using validated instruments to establish prevalence, risk factors, and outcomes essential for evidence-based screening protocols and regulatory frameworks.

XSL•FO
RenderX

Acknowledgments

The author thanks clinic staff at J-Shalom Hospital for facilitating data collection.

Generative artificial intelligence was not used in the writing, analysis, or preparation of this manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

Anonymized data available upon reasonable request to the corresponding author.

Authors' Contributions

Conceptualization: SL
Data collection: SL
Formal analysis: SL
Writing – original draft: SL

Conflicts of Interest

None declared.

References

1. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. Int J Adolesc Youth 2020 Dec 31;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]
2. Wang YY, Wang YS. Development and validation of an artificial intelligence anxiety scale: an initial application in predicting motivated learning behavior. Interactive Learning Environments 2022 Apr 3;30(4):619-634. [doi: 10.1080/10494820.2019.1674887]
3. Li an DE. Mental health in the “era” of artificial intelligence: technostress and the perceived impact on anxiety and depressive disorders-an SEM analysis. Front Psychol 2025;16:1600013. [doi: 10.3389/fpsyg.2025.1600013] [Medline: 40524827]
4. Stănescu DF, Roma canu MC. AI anxiety and neuroticism in attitudes toward AI. Eur J Sustain Dev 2024;13(4):191. [doi: 10.14207/ejsd.2024.v13n4p191]
5. Wang X, Zhou Y, Zhou G. The application and ethical implication of generative AI in mental health: systematic review. JMIR Ment Health 2025 Jun 27;12:e70610. [doi: 10.2196/70610] [Medline: 40577783]
6. Stokel-Walker C. AI driven psychosis and suicide are on the rise, but what happens if we turn the chatbots off? BMJ 2025 Oct 24;391:r2239. [doi: 10.1136/bmj.r2239] [Medline: 41136240]
7. Campbell LO, Babb K, Lambie GW, Hayes BG. An examination of generative AI response to suicide inquires: content analysis. JMIR Ment Health 2025 Aug 14;12:e73623. [doi: 10.2196/73623] [Medline: 40811811]
8. Rahsepar Meadi M, Sillekens T, Metselaar S, van Balkom A, Bernstein J, Batelaan N. Exploring the ethical challenges of conversational AI in mental health care: scoping review. JMIR Ment Health 2025 Feb 21;12:e60432. [doi: 10.2196/60432] [Medline: 39983102]
9. National Code of Health Research Ethics: Federal Ministry of Health; 2007. URL: https://healthresearchwebafrica.org.za/files/NCHRECurrentVersion.pdf [accessed 2026-07-21]
10. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998 Apr 30;17(8):857-872. [doi: 10.1002/(sici)1097-0258(19980430)17:8<857::aid-sim777>3.0.co;2-e] [Medline: 9595616]

Abbreviations

AI: Artificial Intelligence
AIAS: Artificial Intelligence Anxiety Scale
CI: Confidence Interval
C-SSRS: Columbia–Suicide Severity Rating Scale
NCHRE: National Code of Health Research Ethics

XSL•FO
RenderX

Edited by E Mensah; submitted 24.Sep.2025; peer-reviewed by A AL-Asadi, M Salim, O Oyetunji; revised version received 11.May.2026; accepted 11.May.2026; published 29.Jul.2026.

Please cite as:
Lindo S
AI-Related Stress and Suicidal Tendencies in a Primary Care Cohort: Clinical Audit
Online J Public Health Inform 2026;18:e84661
URL: https://ojphi.jmir.org/2026/1/e84661
doi:10.2196/84661

© Shemarah Lindo. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 29.Jul.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Use of Tomographs in Brazil's National Health System: Case Study on the Efficiency of the Public Network in Rio Grande do Norte

Tiago de Oliveira Barreto[1], PhD; Elinaldo Bernardo de Oliveira Júnior[1], BSc; Israel José dos Santos Felipe[1], PhD; Jordana Crislayne de Lima Paiva[1], MD; Nicolas Vinicius Rodrigues Veras[1], PhD; Lorena de Macedo Silva[1], BSc; João Maria Macedo da Costa[1], BSc; Érika Santos de Aragão[2], PhD; Thaísa Góis Farias de Moura Santos Lima[1], PhD; Célio da Costa Barros[3], MD; Rafael Silveira e Silva[4], PhD; Fernando Rocha de Andrade[5], MD; Pablo Holanda Cardoso[1], PhD; Andréa Santos Pinheiro[1], MD; Guilherme Medeiros Machado[6], PhD; Claudia Maria Miranda Veloso[7], PhD; Luciana Simões Camara Leão[8], MSc; Karilany Dantas Coutinho[1], PhD; Ricardo Alexsandro de Medeiros Valentim[1], PhD

[1]Laboratory of Technological Innovation in Health (LAIS), Federal University of Rio Grande do Norte (UFRN), Av. Nilo Peçanha, 620 - Petrópolis, Natal, Brazil

[2]Institute of Public Health (ISC), Federal University of Bahia (UFBA), Salvador, Brazil

[3]Specialized Health Audit Unit, Federal Court of Auditors, Natal, Brazil

[4]Brazilian Institute of Education, Development and Research (IDP), Federal Court of Auditors, Brasília, Brazil

[5]Public Prosecutor's Office in Rio Grande do Norte (PGR), Federal Public Ministry, Natal, Brazil

[6]LyRIDS, ECE-Engineering School, Paris, France

[7]Research Unit in Governance, Competitiveness and Public Polices (GOVCOPP), University of Aveiro, Center for Global Studies (ESTGA), Aveiro, Portugal

[8]Secretariat of Science, Technology and Innovation, Ministry of Health, Brasília, Brazil

Corresponding Author:
Tiago de Oliveira Barreto, PhD
Laboratory of Technological Innovation in Health (LAIS), Federal University of Rio Grande do Norte (UFRN), Av. Nilo Peçanha, 620 - Petrópolis, Natal, Brazil

Abstract

Background: The evaluation of health technologies is fundamental to the sustainability of the Brazilian National Health System (SUS), especially in highly complex and costly procedures such as computed tomography (CT).

Objective: This study analyzed the financial efficiency of CT scanner use in the SUS in the state of Rio Grande do Norte (Brazil), considering the distribution of supply, the profile of providers, and the idle capacity of the public network.

Methods: A total of 49,061 CT scan requests registered in the state system "RegulaRN Ambulatorial" between October 2023 and January 2025 were examined. Subsequently, data cleaning and grouping were performed.

Results: The results indicated that 66.8% (24,769/37,089) of requests were from patients with cancer, 61.2% (22,703/37,089) were women, predominantly from municipalities with higher population densities, such as Natal and Mossoró, with an average waiting time of 44.86 (SD 88.43) days. More than 36,000 (99%) requests for the examinations were performed by private and philanthropic hospitals contracted by SUS, while public units with CT scanners accounted for less than 1% of production, evidencing significant underutilization.

Conclusions: The findings suggest that strategically shifting demand to the public health system can increase efficiency, reduce costs and waiting times, and promote greater equity in access, contributing empirical evidence for more rational public policies on the use of diagnostic technologies.

(Online J Public Health Inform 2026;18:e83494) doi:10.2196/83494

KEYWORDS

health economics; RegulaRN Ambulatorial; computed tomography; health regulation; public health

XSL•FO
RenderX

Introduction

Health Economics in the Context of Tomographs in Rio Grande do Norte

Efficiency in public health management is a central and recurring theme in discussions about the sustainability of health systems and the quality of services offered to the population. In Brazil, the Brazilian National Health System (SUS) represents a universal and free model, designed to ensure the constitutional right to health. However, this system faces significant challenges due to the scarcity of financial resources and the constant growth in demand for services, circumstances that intensify the need for rational and efficient allocation of available resources to ensure equity in access and the quality of care provided [1-5].

Within the context of health economics, analyzing the costs related to the state apparatus, as well as monitoring the investment scenario, is essential to promoting efficiency in access to public health [6]. Due to Brazil's continental size (approximately 8,510,000 km²) and significant population (more than 212 million inhabitants), the country invests billions of Brazilian reais in public health, approximately 3.88% of the federal budget, which leads us to question whether the amount is poorly distributed [7]. In the context of Latin American health systems, Brazil faces persistent challenges in financing public health, with a significant portion of the population depending on services provided by the private sector, despite the universal coverage proposed by SUS [7-9].

Promoting health monitoring through digital health systems can help to address the greatest weaknesses in the public health system [10,11]. One of the important topics for monitoring is the performance of consultations and examinations, especially highly complex examinations. In this sense, computed tomography (CT) is a fundamental clinical procedure for the diagnosis of various diseases, constituting an essential step in many care flows [12]. However, the uneven distribution of this equipment and the variation in the degree of utilization among service providers can compromise both access and the efficiency of the system. To this end, health systems increasingly need timely and efficient regulatory systems.

Recent Brazilian studies highlight regional disparities in the supply and the use of CT scanners in the SUS, as well as pointing out challenges related to production capacity and the allocation of technological resources. Research such as that conducted by Santos [13] and dos Santos et al [14] discusses the availability and degree of the use of CT scanners, whereas investigations carried out by Dovales et al [15] and Viacava et al [16] address issues related to the distribution of resources and the quality of services, which reinforce the urgent need to optimize the use of this equipment in the face of budget constraints and constantly increasing demand.

RegulaRN Ambulatorial is the current system responsible for managing outpatient health care services in Rio Grande do Norte. This digital health system is responsible for coordinating and systematizing requests for tests, while organizing the queues for procedures offered by the SUS. The RegulaRN Ambulatory system incorporates an ecosystem of digital health solutions, unlike other health regulation platforms such as SISREG (National Regulation System) and SIGUS (Outpatient Regulation System), which present problems of interoperability and transparency of information. This paper analyzes requests for CT exams, investigating the execution of this type of procedure for the evaluation of the efficiency of public resource use, in addition to identifying possible regional disparities, efficiency, and effectiveness in the procedures and services provided [17-19].

In the field of health economics, efficiency refers to the ability of a health system to maximize the use of available resources to deliver the greatest possible volume of services and health outcomes. In this context, cost-efficiency is related to the capacity to provide services at lower public expenditure, while maintaining adequate access and quality of care. Evaluating these aspects becomes particularly relevant in the case of high-cost diagnostic technologies such as CT, which require significant financial investment and specialized infrastructure [1].

Given this scenario, the study begins with the following question: what does the analysis of CT scanner utilization and associated public expenditure in the SUS in Rio Grande do Norte reveal: efficiency or waste? In this context, efficiency is understood as the use of equipment to maximize access and clinical outcomes with available resources. Waste, on the other hand, occurs when there is underutilization or inadequate allocation of equipment and resources. The objective is to analyze the use of CT scanners in the public health system of Rio Grande do Norte and to assess the effectiveness of the service, the distribution of supply, the profile of providers, and the adequacy of allocated resources. The relevance of the research lies in providing support for improving the efficiency of public health management in the state of Rio Grande do Norte and proposing strategies that can be replicated in other national or international contexts.

Sociodemographics and Regional Health Mapping of Rio Grande do Norte

Rio Grande do Norte is a state located in the Northeast region of Brazil, characterized by its socioeconomic and demographic diversity. According to data from the Brazilian Institute of Geography and Statistics [20], this state has approximately 3.3 million inhabitants spread across 167 municipalities and occupies a land area of 52,811 km². The capital, Natal, stands out as the most populous municipality, with over 750,000 inhabitants, followed by Mossoró, with around 264,000 inhabitants, and Parnamirim, with approximately 252,000 residents. The urbanization rate in Rio Grande do Norte is high, reaching 77.8%, while the average population density is 62.54 (DP 34) inhabitants per square kilometer, concentrated in coastal regions.

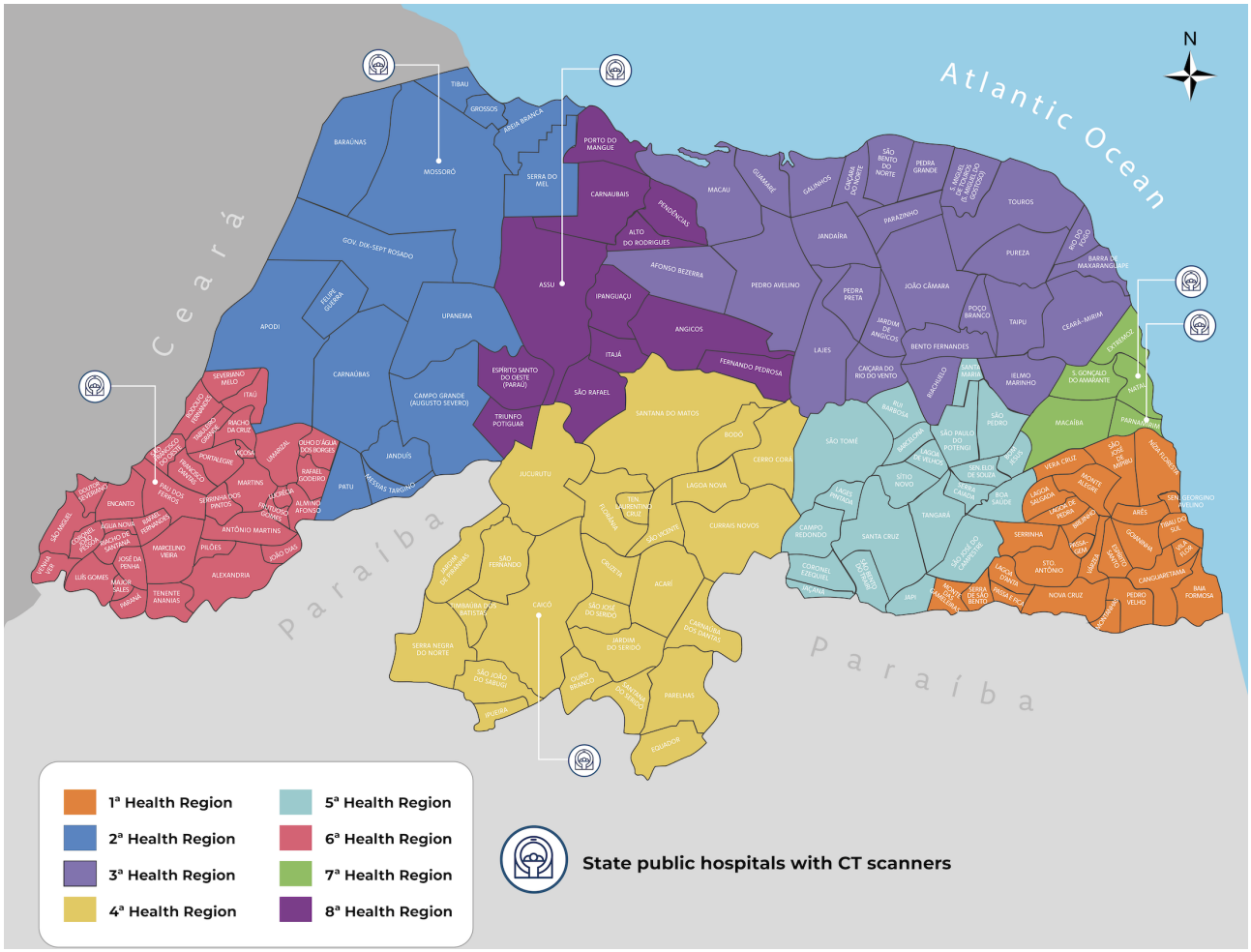
In the context of public health, Rio Grande do Norte is currently divided into 8 Health Regions, named in an ordinary manner. Each region has a municipality with greater representation, responsible for managing the health regulation process in the surrounding municipalities. In the first Health Region, the hub municipality is São José de Mipibu, followed in order by Mossoró, João Câmara, Caicó, Santa Cruz, Pau dos Ferros,

XSL•FO
RenderX

Natal, and finally, representing the eighth, Assu [21,22]. Figure 1 shows the distribution of Health Regions.

Furthermore, according to data from the National Register of Health Establishments [23] and the government data platform TabNET [24], Rio Grande do Norte has 6173 health units. Of these, 1326 are basic health units under municipal management, 10 are emergency care units, and 21 are state hospitals. Of the hospitals in the state network, 8 are concentrated in the seventh region, 5 in the second region, 2 in the first and fourth regions, and 1 in the third, fifth, sixth, and eighth regions. This configuration reflects the centralization of hospital services in the seventh (Natal) and second (Mossoró) regions, which may influence access to and the provision of specialized services to the population [21,22].

Figure 1. Distribution of Health Regions in Rio Grande do Norte. CT: computed tomography.



Methods

Characterization of the Research Method

This study is characterized as applied research, with a quantitative, descriptive, and exploratory approach, focused on analyzing the use of CT scanners in the state of Rio Grande do Norte. The investigation identifies the distribution of care between public sector health units and contracted private units, considering the potential absorption of demands currently met in both segments, with a focus on rationalization and identification of possible bottlenecks.

The data analyzed were extracted from the RegulaRN Ambulatorial platform, a state system for regulating SUS procedures in Rio Grande do Norte. The database includes requests for CT scans registered between October 2023 and January 2025, with a total of 49,061 records. Each record initially contained 16 variables, which included information about the request, service, provider, and administrative characteristics of the procedure. It should be noted that the platform was structured to migrate data from the platform previously used to regulate state exams. However, these data were not considered in this study because they were not coherent with the study's objectives.

The RegulaRN Ambulatory system records requests for specialized procedures submitted by health care professionals within the public health network. Each record corresponds to a request for a diagnostic procedure and is associated with an individual patient. However, the same patient may generate multiple requests over time depending on clinical needs, and each request corresponds to a specific examination or procedure. For this study, all CT scan requests registered in the system during the analyzed period were extracted and organized for analysis.

The methodological process was structured in four sequential stages: (1) extraction, in which the original columns from the database were obtained, namely request date, priority, status,

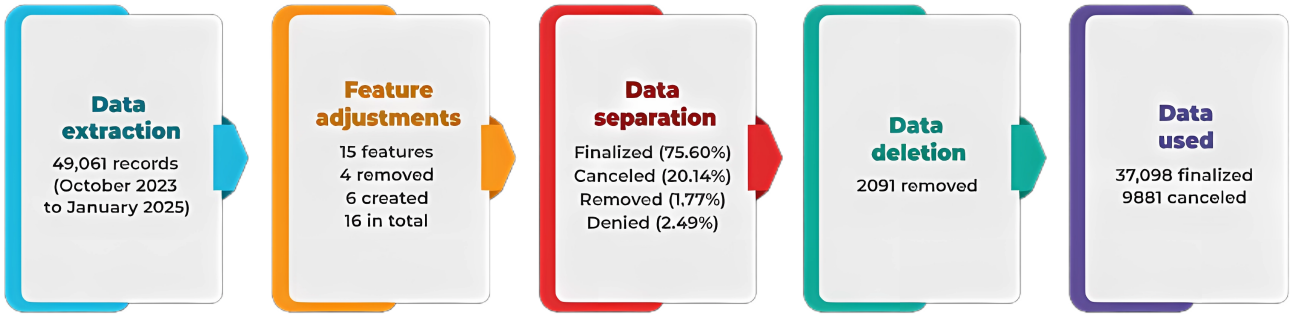
XSL•FO
RenderX

patient name, gender, requesting unit, unit municipality, procedure name, subgroup, APAC (Outpatient Health Procedure Authorization) number, authorization date, regulating physician, provider unit, provider municipality, date of care, and date of completion. Following this, we carried out stage (2), feature suitability, in which we preprocessed the temporal variables (dates), which initially had a unified date and time value. Therefore, we separated the date and time and then created derived variables representing the time intervals between the following steps: request and service, request and authorization, authorization and service, and service and closure. In addition, some characteristics were removed, namely columns that contained sensitive variables or that would not contribute significantly to this study, such as the patient's name, procedure subgroup, APAC number, and name of the regulating physician.

It was considered a separation stage (3), given that the status of the requests could contain 4 different types of outcomes, namely completed (37,089/49,061, 75.60% registers), canceled (9881/49,061, 20.14% registers), removed (868/49,061, 1.77% registers), and denied (1222/49,061, 2.49% registers). Thus, this analysis focused on records with completed and canceled status, as they presented greater volume and relevance to the study objectives. Records with removed and denied status were disregarded because they were not significant. Finally, an exclusion stage (4) was included, in which data from outcome variables with removed and denied status were removed, as they did not contribute to the evaluation and objectives of this study. In this regard, the data used for this study consisted of 37,089 completed requests and 9881 canceled requests. Figure 2 shows the sequence of this methodological process.

In order to enable the reproducibility of the methodological process, the anonymized data used to compose this research, within the specified period, can be accessed via the link on Zenodo [25], developed by CERN (European Organization for Nuclear Research), openly and free of charge.

Figure 2. Pipeline created for curating computed tomography exam data from the RegulaRN Ambulatorial system.



Ethical Considerations

The requirement for ethical approval was waived by the Research Ethics Committee of the Federal University of Rio Grande do Norte for the studies involving humans because according to Resolution 674/2022 of the National Health Council of the Ministry of Health [26], this research is exempt from registration with the Research Ethics Committee/Brazil or the National Research Ethics Commission (CONEP)/Brazil because, according to the regulations, aggregated and anonymized data do not require an ethics committee review. The studies were conducted in accordance with the local legislation and institutional requirements.

Results

The results of this survey were organized to consider completed requests and canceled requests separately, in order to provide greater clarity and detail in the analyses presented.

Requests Completed

When considering the data profile of RegulaRN Ambulatorial, from October 2023 to January 2025, the 37,089 records of exam requests with the status "Completed" show a greater occurrence of data from patients with cancer (in terms of priority), females (in terms of gender), requested mainly by 2 municipalities in the state of Rio Grande do Norte, Natal and Mossoró, for CT of the chest, upper abdomen, and skull (in terms of procedure). Table 1 clearly shows the distribution of request data with the status "Completed."

XSL•FO
RenderX

Table . Distribution of data from completed requests (N=37,089).

Feature	Values, n (%)
Categorical values	
Priority	
Oncology	24,769 (66.8)
Hospitalized	2221 (6)
Very high	2654 (7.2)
High	3831 (10.3)
Medium	2999 (8.1)
Low	601 (1.6)
Lawsuit	14 (<0.1)
Sex	
Male	14,386 (38.8)
Female	22,703 (61.2)
Requesting unit	
Health unit 174	8336 (22.5)
Health unit 23	2966 (8)
Health unit 4	1622 (4.4)
Health unit 20	1585 (4.3)
Health unit 19	1535 (4.1)
Others	21,045 (56.7)
Requesting municipality	
Natal	10,815 (29.2)
Mossoró	3691 (10)
Caicó	1133 (3.1)
Others	21,450 (57.8)
Procedure	
Chest CT[a] scan	9815 (26.5)
Upper abdominal CT scan	9688 (26.1)
Skull CT scan	9158 (24.7)
Others	8428 (22.7)
Service provider unit	
Health unit 4	20,258 (54.6)
Health unit 3	6388 (17.2)
Health unit 12	5264 (14.2)
Health unit 11	1608 (4.3)
Health unit 2	1466 (4)
Others	2105 (5.7)
Municipality provider	
Natal	27,527 (74.2)
Caicó	1621(4.4)
Numerical values	
Request to service	
<30 days	25,256 (68.1)

XSL•FO
RenderX

Feature	Values, n (%)
>30 days	11,833 (31.9)
Request to authorization	
<30 days	27,476 (74.1)
>30 days	9613 (25.9)
Authorization to service	
<30 days	35,716 (96.3)
>30 days	1373 (3.7)
Service to closure	
<30 days	31,512 (85)
>30 days	5577 (15)

[a]CT: computed tomography.

Regarding time indicators, the average interval between the request and the service was 44.86 days, with a standard deviation of 88.43 days and a median of 20 days (IQR 8-40). The period between the request and authorization had an average of 38 days (SD 86.88 days) and a median of 12 days (IQR 2-32). The interval between authorization and treatment was, on average, 7 days (SD 15.51 days), with a median of 2 days (IQR 1-8). Finally, the average time between admission and discharge was 22.15 days, with a standard deviation of 30.91 days and a median of 14 days (IQR 8-22). These results indicate that the longest delays occur during the authorization and service periods, suggesting that demand exceeds the health care network's capacity to respond promptly.

Another relevant aspect concerns the distribution of agreements for the provision of examinations. The metropolitan areas of Natal, Mossoró, and Caicó are hubs of reference for other municipalities. Because they have more robust infrastructure, these hub municipalities respond, through agreements, to requests from patients in other municipalities, whether or not those municipalities are located in their Health Region. Thus, Table 2 shows the number of completed requests from municipalities that performed some type of CT scan under agreement with the reference centers.

In addition, it was possible to identify significant patient transfers between regions that either have or do not have health care facilities equipped with CT scanners. Table 3 presents data on the 3 provider units that receive the most requests and the municipalities and Health Regions from which the requests originate. It was possible to identify that health unit 4, belonging to the seventh Health Region of Rio Grande do Norte, expressly serves not only residents of its region but also those of other surrounding regions.

XSL•FO
RenderX

Table . Distribution of completed requests to service provider units.

Municipality provider	Requesting municipality, n (%)
Natal (n=27,527)	
Natal	10,810 (39.3)
São Gonçalo do Amarante	1102 (4)
Macaíba	1056 (3.8)
Parnamirim	967 (3.5)
Ceará-Mirim	664 (2.4)
Caicó	614 (2.2)
Others	12,314 (44.7)
Mossoró (n=7941)	
Mossoró	3650 (46)
Açu	570 (7.2)
Apodi	367 (4.6)
Baraúna	304 (3.8)
Caraúbas	232 (2.9)
Areia Branca	173 (2.2)
Others	2645 (33.3)
Caicó (n=1621)	
Caicó	519 (32)
Parelhas	163 (10.1)
Currais Novos	134 (8.3)
Apodi	108 (6.7)
Jucurutu	77 (4.8)
Serra Negra do Norte	66 (4)
Others	554 (34.2)

XSL•FO
RenderX

Table . Distribution of the 3 service providers that receive the most requests and the municipalities/regions from which they are requested.

Service provider unit (Health Region)	Municipality	Municipal Health Region
Health unit 4 (7th)		
Natal	7304	7th
Parnamirim	922	7th
Macaíba	875	7th
São Gonçalo do Amarante	817	7th
Caicó	600	4th
Ceará-Mirim	534	3rd
Santa-Cruz	437	5th
Touros	308	3rd
São José de Mipibu	286	1st
Nova Cruz	275	1st
Health unit 3 (2nd)		
Mossoró	3301	2nd
Apodi	209	2nd
Açu	245	8th
Baraúna	222	2nd
Caraúbas	204	2nd
Areia Branca	121	2nd
Patu	113	2nd
São Miguel	89	6th
Carnaubais	82	8th
Grossos	82	2nd
Health unit 12 (7th)		
Natal	2830	7th
São Gonçalo do Amarante	216	7th
Ceará-Mirim	99	3rd
Macau	95	3rd
Nova Cruz	89	1st
Goianinha	68	1st
Touros	62	3rd
Canguaretama	62	1st
Guamaré	59	3rd
Pedro Avelino	58	3rd

The procedures performed by health care providers are also relevant elements for analysis, especially those identified as the most frequently performed in the respective health care facilities. According to Table 1, health unit 4, health unit 3, and health unit 12 are among the institutions that receive the most patients. It is worth noting that the first 2 units mentioned are references for the care of patients with cancer, representing approximately 66.8% of completed requests. Thus, in a preliminary assessment, it was already expected that these units would have a higher number of requests. In addition, as shown in Table 1, there are 3 tests that are in high demand compared to the others, namely CT scan of the upper abdomen and CT scan of the pelvis, lower abdomen, and chest. Therefore, it is expected that these procedures will also be widely demanded among these units (see Table 4 in Multimedia Appendix 1).

Still, when analyzing which units perform the most different types of procedures, health unit 4 was responsible for performing the most upper abdomen CT scans (5362 in 2024), thoracic spine with or without contrast (81 in 2024), pelvis/lower abdomen (4899 in 2024), face/sinuses/temporomandibular joints (361 in 2014), chest (5790 in 2024), neck (833 in 2024), sella turcica (22 in 2024), and positron emission tomography

XSL•FO
RenderX

(PET)-CT (367 in 2024). Health unit 12 performed the highest number of CT scans of the upper limb (26 in 2024) and lower limb (74 in 2024), lumbosacral spine with or without contrast (314 in 2024), and skull (1183 in 2024). Finally, health unit 3 performed the most CT scans of the cervical spine with and without contrast (88 in 2024). A highlight of this analysis is that, in theory, any provider could perform any of the tomography exams, except for PET-CT, which requires a higher level of technology. However, the data show a high distribution in terms of the number of inmates across different units. In addition, not all procedures refer to all units.

We consider analyzing the relationship between the priority of care, the provider unit, and the municipality of the unit that requests the most (Table 5 in Multimedia Appendix 1). Our argument is that it is necessary to identify whether there is any relation or direction between a specific service profile for a specific unit and whether the number of requests for that unit is coming from a nearby municipality or not.

The study found that health units 3 and 4, although they are oncology centers, also address other priorities that could be allocated to other units, which contribute to a better distribution of demand. On the other hand, public hospitals, such as health unit 5 (13 requests) and health unit 9 (87 requests), showed low service levels, with few records, most of them from outside municipalities, suggesting potential for a greater use of this equipment and a reduction in waiting times.

Another important element to highlight is the assessment of which CT scans are most frequently requested over time, as well as their cost, given that the most requested ones do not necessarily incur higher costs for the state government. This will enable the health care network to gain a general understanding of which specific tests are most in demand and, as the network expands, what adjustments need to be made.

According to the data obtained from the RegulaRN Ambulatorial database for 2023 (as of October), CT of the upper abdomen, CT of the pelvis and upper abdomen, and CT of the thorax cost approximately R$ 19,130.94, R$ 18,853.68, and R$ 19,779.45, respectively (Table 4; a currency exchange rate of R$ 1=US $0.199 is applicable). For the year 2024, the most costly and in-demand procedures were also CT of the upper abdomen, CT of the pelvis and upper abdomen, and CT of the thorax, with the significant emergence of PET-CT, costing approximately R$ 1,239,629.46, R$ 1,172,809.80, R$ 1,233,282.81, and R$ 840,780.78, respectively. In January 2025, the same exams remained the most in demand, with costs of R$ 84,287.04 (upper abdomen CT scan), R$ 77,910.06 (pelvis and lower abdomen CT scan), R$ 85,801.89 (chest CT scan), and R$ 52,680.50 (PET-CT scan).

XSL•FO
RenderX

Table . Representation of the total cost of procedures per year.

Tomography procedure	October 2023 (R$[a])	October 2024 (R$)	January 2025 (R$)
CT[b] scan of upper abdomen (R$ 138.63/exam)	19,044	1233.996	83,904
CT scan of lower limb joints (R$ 86.75/exam)	86.75	17,003	2689.25
CT scan of upper limb joints (R$ 86.75/exam)	86.75	5465.25	347
CT scan of the cervical spine with or without contrast (R$ 86.76/exam)	433.8	26,114.76	694.08
CT scan of the lumbosacral spine with or without contrast (R$ 101.10/exam)	1415.4	70,871.1	2628.6
CT scan of the thoracic spine with or without contrast (R$ 86.76/exam)	347.04	21,776.76	1041.12
CT scan of the face/sinuses/temporomandibular joints (R$ 86.75/exam)	1648.25	81,631.75	6072.5
CT scan of the pelvis/lower abdomen (R$ 138.63/exam)	18,853.68	1,172,809.8	77,910.06
CT scan of appendicular segments (arm, forearm, hand, thigh, leg, and foot) (R$ 86.75/exam)	173.5	5638.75	347
CT scan of the sella turcica (R$ 97.44/exam)	0.00	5066.88	292.32
CT scan of the thorax (R$ 136.41/exam)	19,779.45	1,233,282.81	85,801.89
CT scan of the skull (R$ 97.44/exam)	6138.72	368,323.2	16,662.24
CT scan of the neck (R$ 86.75/exam)	1735	96,032.25	7200.25
Positron emission tomography (PET-CT) (R$ 2107.22/exam)	8428.88	840,780.78	52,680.5
Total	78,170.54	5,178,793.09	338,270.56

[a]A currency exchange rate of R$ 1=US $0.199 is applicable.

[b]CT: computed tomography.

Regarding the costs associated with performing CT scans, considering that there are public and private health care facilities contracted and paid for with public funds, it is necessary to analyze the distribution of resources used in public health care throughout the period of this study. Finally, analyzing the budget distribution between public and private entities under contract, it was observed that public units (health units 5, 6, 8, and 9) accounted for less than 1% of the total investment in CT scans, with most resources directed to private and philanthropic health care networks (units 1, 2, 3, 4, 7, 10, 11, 12, and 13).

The data show that as of October 2023, the budget for CT scans was R$ 73,170.54 (100%), with the entire amount spent in the private sector. From January to December 2024, R$ 45,366.42 (0.9%) of the cost of the exams was spent in the public sector and R$ 5,133,436.67 (99.1%) in the private sector. Concluding with January 2025, R$ 2,731.74 (0.8%) was spent in the public sector and R$ 335,538.82 (99.2%) in the private sector.

Canceled Requests

A total of 9881 records of canceled requests were analyzed, the evaluation of which is essential for identifying patterns and possible gaps in the regulatory process. Cancelation may occur at any time, from the request until after service, and is carried out by the requesting unit. Table 8 in Multimedia Appendix 1 details the distribution of cancelations and indicates that the highest concentration occurs in the initial stages of the request, as well as after the service has been provided by the provider unit. This pattern suggests inefficiencies in the early identification of requests that will be canceled, resulting in the unnecessary expenditure of time and human resources in analyzing requests that could have been discarded beforehand.

Following the results, it was also possible to identify that among the canceled requests, there is a greater significance in requests prioritizing patients with cancer and female patients, which was expected given the greater significance throughout the dataset. As for the requesting unit and procedure indicators, there are

XSL•FO
RenderX

the following variations: among the canceled requests, the highest number of requests comes from health unit 102, belonging to the municipality of Macaíba, which is located in the metropolitan region of Natal. Although this municipality does not appear significantly in the number of completed requests, it has a higher number of requests canceled in the initial phase. This may represent some difficulty that this municipality faces or has faced in making new requests.

As for the procedure, CT of the upper abdomen, which is one of the most requested examinations, is also among the most canceled. For requests that were only authorized, it should be noted that the requesting units are hospital units, with health unit 32 having the highest recurrence of cancelations after authorization. The most canceled procedure at this stage is chest CT. As for requests that were actually fulfilled, health unit 23, belonging to the municipality of Natal, has the highest number of cancelations. The most canceled procedure at this stage is chest CT.

Discussion

The discussion of the results obtained in this paper will be conducted from the perspective of the Brazilian Constitution and the outcomes achieved. However, the results obtained indicate the underutilization of public resources in favor of their use in private institutions.

The Optimization of Health Services as a Constitutional Obligation

The 1988 Constitution of the Federative Republic of Brazil, by establishing the SUS and adopting a model of universal and free access inspired by the British system (Beveridge model), recognized health as a right for all and a duty of the State. Under Article 198 of the Constitution, public health actions and services must form a regionalized and hierarchical network, with decentralized organization and participation by federal entities—the Union, states, Federal District, and municipalities—each of which is responsible for adopting legal, economic, and practical measures aimed at realizing this fundamental right [27].

This constitutional effort to structure the SUS as a federation highlights the joint responsibility of the entities involved in ensuring the provision of quality health services to the population. To this end, public managers are required to perform their duties as efficiently as possible to optimize the relationship between revenue and expenditure. This guideline is reinforced by the principle of efficiency, enshrined in the heading of Article 37 of the Federal Constitution, which should guide all administrative activity.

The existence of essential health equipment and services in a state of idleness is incompatible with the constitutional order and governing legislation, especially given the growing demands of the Brazilian population. This incompatibility becomes even more evident in the case under review when it is noted that, although there is ordinary and continuous demand for CT scans, this demand has been met by the private sector as if it were an exceptional service, while a public network capable of offering such procedures remains idle.

Law number 8,080/1990, which regulates the SUS, establishes in Article 7, items II, IV, and XI that health actions and services must be organized based on the principles of comprehensive care, preservation of personal autonomy, and rational use of resources and that it is the duty of managers to promote the full use of installed capacity, avoid waste, and ensure the efficient use of available resources [28].

The regulation of public health financing, in turn, was consolidated with Complementary Law number 141, of January 13, 2012, which regulates paragraph 3 of Article 198 of the Constitution. The law defined the criteria for apportioning federal, state, and municipal resources, defined what is meant by expenditures on public health actions and services (Article 3), and reaffirmed the minimum revenue requirements set forth in Article 77 of the Brazilian Act of Transitory Constitutional Provisions (ADCT). In addition, Complementary Law 141/2012 reinforced the mandatory use of health funds and emphasized the need for transparent mechanisms for controlling and monitoring public spending (Articles 12, 14, and 16) [29].

Constitutional Amendment Number 86/2015 (known as the “Mandatory Budget Amendment”) reinforced this regulatory framework by establishing a minimum level of federal health spending in the Constitution, making it mandatory to allocate a portion of individual parliamentary amendments to the health sector. The amendment also linked funds from oil royalties to public health actions and services, with the aim of increasing funding and ensuring its stability. In turn, Constitutional Amendment Number 95 of 2016, which was only reversed in 2023, established a freeze on public spending for 20 years, imposing an environment of great fiscal austerity, in which health was heavily impacted by the loss of resources [30].

Given the regulatory landscape of fiscal austerity and the recognized budgetary constraints that limit government action, efficiency in public health management is a constitutional duty, directly linked to the state’s obligation to ensure the fundamental right to health in a universal, equitable, and comprehensive manner. The SUS is facing growing demands and limited resources, so public managers need to make the best use of available resources, getting the most out of public policies and avoiding waste.

In this scenario, the adoption of management practices aimed at rationalizing expenses and enhancing installed capacity becomes imperative. In this case, it was found that, even in the face of continuous and routine demand for CT scans, the State chose to pay for such services from the private network, despite the existence of idle public equipment that was suitable and available to meet the same demand.

Such conduct constitutes flagrant administrative inefficiency and disregard for the constitutional principles of economy, legality, and good governance. Discussing health economics, therefore, is not just a management issue but a legal and ethical requirement, essential for addressing social and regional inequalities. Article 6 of the Federal Constitution establishes health as a fundamental social right, whereas Articles 196 et seq impose on the State the duty to guarantee it based on the principles of universality, equity, and comprehensiveness [2].

XSL•FO
RenderX

Thus, the constitutional order establishes the imperative that public health management acts efficiently, especially in the use of highly complex technologies such as CT, promoting the best use of the existing public infrastructure before resorting, unjustifiably, to the private sector.

Therefore, the search for the optimal path in economic terms, especially with regard to the allocation of highly complex equipment such as CT scanners, is directly linked to the legal duty to promote strategic planning, the evaluation of health technologies, and the maximization of the installed capacity of the public system. The existence of tomography equipment lying idle in public health facilities, when there is demand for the service, not only represents a waste of resources but also constitutes a flagrant affront to the constitutional principles of efficiency, economy, and reasonableness in public administration.

In the specific case of Rio Grande do Norte, the severity of the lack of availability of CT scan services is even more pronounced. As shown in this study, the demand for CT scans is concrete, growing, and well-documented, but it has been met mainly by private providers contracted by the SUS, even in regions where public CT scanners exist. However, the interpretation of underutilization should also consider possible operational constraints, such as equipment maintenance status, workforce availability, and local service organization, which were not directly measured in this study. This finding may indicate operational challenges in the organization of CT services, raising questions about the efficiency of resource allocation and the optimization of existing public infrastructure, which impose on the public administration the duty to optimize resource management—that is, to use the least amount of public resources possible to achieve the highest degree of effectiveness in the service provided.

The allocation of public funds for the contracting of private tomography services, to the detriment of the use of existing public infrastructure, demonstrates a failure to comply with the duty of management optimization, constituting conduct that is uneconomical, inefficient, and inconsistent with the provisions of Article 37 of the Federal Constitution. The allocation of public funds to contract private CT services, while public equipment remains underutilized, suggests potential inefficiencies in resource allocation and highlights the need for improved coordination between available infrastructure and service demand.

The Inefficiency of Health Care in Rio Grande do Norte: Discussion of the Data Obtained

Considering the scarcity of financial resources within the SUS, efficiency in public management is an essential element for the sustainability of the system. Historically, the public health sector in Brazil has had high health care costs, which can be attributed to multiple factors, such as increased demand, the incorporation of advanced technologies, a shortage of qualified professionals, waste in the health care chain, and institutional management failures. In the case of highly complex exams, such as CT scans, for example, which involve higher public funding costs, monetary management must be more effective [31,32].

The findings of this study reveal that most patients undergoing CT scans through the SUS in the state of Rio Grande do Norte are primarily patients with cancer, with a significant concentration of care in 3 contracted private health care facilities. In view of this, a relevant discussion concerns the possibility of distributing these services to other hospitals, mainly in state hospitals, which, according to the data, have a low number of patients or promote a greater number of consultations for patients without cancer among state hospitals.

It should be noted that, of the 21 state hospitals, 8 of them have CT scan equipment to meet the demand, namely Giselda Trigueiro Hospital, Dr Tarcísio Maia Regional Hospital, Dr Cleodon Carlos de Andrade Hospital, Nelson Inácio dos Santos Regional Hospital, Telecila Freitas Fontes State Hospital, Coronel Pedro Germano Central Hospital, Deoclécio Marques de Lucena Regional Hospital, and Monsenhor Walfredo Gurgel Hospital. However, only 3 of these hospitals performed some type of CT scan to meet the state's outpatient care demand, and even then, they performed a relatively small number of scans when compared to other units.

According to studies available in academic literature, the minimum capacity of a CT scanner is around 40 exams per day [33,34]. Therefore, the hospitals mentioned above are far below their maximum operational capacity. From this perspective, it is suggested that competent institutions promote a protocol for redistributing the allocation of demands by at least 20% of what is suggested in the literature, not only to reduce public spending but also to reduce average waiting times.

The need to increase the number of appointments at state hospitals in the network is essential to promote greater patient care capacity and minimize waiting times [35]. However, for this increase to be feasible, it is necessary to expand the available human resources, as well as invest in the training and development of the teams, in order to guarantee the quality and efficiency of the service. Given that there is a continuous flow of patients with cancer, those with lower priority in the queue wait considerably longer to be seen to the point that they choose to have the examination conducted at a private facility at their own expense.

From this perspective, the results indicate 2 important points. The first concerns the fact that, although CT scans are considered a high-cost procedure, there are few records of lawsuits, that is, when the patient initiates a procedure with the justice system to make the examination possible. This phenomenon can be explained by the private sector's ability to offer payment alternatives that facilitate access for patients from the public system. The second aspect concerns the impact of long wait times in public services: patients with a history of public health care services choose to undergo examinations in the private health care network, bearing the costs to ensure faster diagnosis.

Furthermore, from a macroregional perspective, some Health Regions, such as the first, third, and fifth, do not have referral hospitals with CT scan equipment, meaning they do not have the capacity to meet local demand. Therefore, it is necessary for residents to travel a considerable distance so that the nearest providers can serve them. Thus, expanding the capacity of local

XSL•FO
RenderX

hospitals to perform these tests could significantly contribute to reducing the need for travel and increasing access.

It was also observed that, among the 13 hospitals that perform CT scans, there is a significant concentration of procedures in a few contracted private units, while the potential idle capacity of the public network appears underutilized according to the available administrative data. This concentration not only leads to increased public costs but also reduces the autonomy of the SUS in providing essential services. As pointed out by Botega et al [36], Brazilian public hospitals often operate below capacity, which reinforces the need for a strategic redistribution of demand to maximize the use of public resources, reduce costs, and foster competition among providers, thereby enhancing the quality and efficiency of care.

As for records of canceled requests, this work presents a considerable volume of requests. In addition, subdividing cancelation periods is also important for identifying where the main bottlenecks lie, given that a request that is made and then canceled for trivial reasons as the days go by systematically contributes to further delays in service. Therefore, institutional intervention is necessary to resolve the main issues presented in the results of this study [37].

Regarding social impact, optimizing the use of CT scanners in public facilities has the potential to expand access to diagnostic exams, reduce waiting times, and minimize waiting times. This effect would be especially relevant for patients classified as lower priorities, who currently experience low service rates in the system. Furthermore, such improvements would not be limited to regions with greater service availability, extending benefits to smaller municipalities and areas lacking dedicated equipment, thus contributing to the promotion of equity in health care access.

The results for Rio Grande do Norte reveal typical challenges faced by public health systems in developing countries, where the concentration of high-cost technologies in a few urban hubs creates territorial inequalities and pressures referral hospitals, resulting in longer waiting times and potential underdiagnosis of diseases. This situation reinforces the importance of strategic planning for the allocation of new equipment, guided by epidemiological needs, population density, and the installed capacity of health units [38,39].

Finally, the results highlight the need to review contracts between the state, municipalities, and service providers, given that some establishments may be serving beyond the necessary demand, while others face difficulties in ensuring adequate access for the population. This review is essential to ensure the equity, efficiency, and sustainability of the public health system in Rio Grande do Norte.

Conclusions

This study highlights significant distortions in the allocation and utilization of CT scanners in the public health system of Rio Grande do Norte, characterized by the concentration of care in a limited number of contracted private units, despite the existence of underutilized equipment in state public hospitals. These findings suggest the possibility of expanding the installed capacity of the SUS, with potential positive impacts on reducing wait times and optimizing operational costs. Furthermore, it is noteworthy that the future use of analytical tools based on artificial intelligence could provide important support for prioritizing exams and redistributing the health care burden more equitably and efficiently.

The evidence produced in this study can support managers in assessing the absorption capacity of currently outsourced demand, while also reinforcing the importance of more strategic regulation that is sensitive to regional disparities, ensuring equal access to CT scans throughout the state. Regarding limitations, we highlight the incompleteness of data for the entire period between 2023 and 2025, as well as the possibility of inconsistencies in information migrated from previous platforms. The availability of complete historical data would allow for a more comprehensive and contextualized analysis of the performance of service providers. Furthermore, the dataset did not include information on clinical outcomes, diagnostic yield, or patient follow-up, which limits the ability to assess the clinical effectiveness of the performed CT scans. Therefore, to mitigate the impacts arising from the transition between systems in the public sector, we recommend the adoption of periodic audits and data quality frameworks, as recommended by the Department of Informatics of the Unified Health System (DATASUS) [40], to preserve the greatest possible integrity of the information.

For future research, we recommend conducting a longitudinal analysis to assess the effects of restructuring regional contracts, as well as investigating the impact of judicialization resulting from the unavailability of tests. Furthermore, the possibility of replicating the methodology adopted in other high-cost procedures expands the potential contribution to strengthening the rational management of the SUS, promoting greater efficiency and equity in the use of available resources.

Acknowledgments

The authors express their gratitude to the Laboratory for Technological Innovation in Health at the Federal University of Rio Grande do Norte, the Federal University of Bahia, the Federal Public Ministries, the Ministry of Health, the Federal Court of Auditors, LyRIDS (ECE-Engineering School), and the University of Aveiro for their support.

Artificial intelligence (AI) tools were used to support the writing of the manuscript, solely to make the text more organized and concise. In addition, the authors are grateful to Bruna Fernandes de Araújo for her meticulous translation of the original text into English. AI tools were used to support the writing of the manuscript, solely to make the text more organized and concise.

XSL•FO
RenderX

Funding

The authors declare that they have received financial support for the research, authorship, and publication of this article. This study was funded by project 127/2022—A Fundação Norte-Rio-Grandense de Pesquisa e Cultura (FUNPEC)/Federal University of Rio Grande do Norte (UFRN)/International Cooperation—(TED 52/2022-MS), Research project applied to international cooperation: health economics from the perspective of public policy evaluation, carried out by the Laboratory of Technological Innovation in Health of the UFRN.

Authors' Contributions

Conceptualization: TOB, EBOJ, IJSF, PHC, JCLP, ESA, TGFMSL, GMM, KDC, RAMV
Data curation: TOB, EBOJ, IJSF, PHC, JCLP, LMS, JMMC, NVRV, FRA, GMM, CMMV
Formal analysis: TOB
Methodology: TOB, EBOJ, IJSF, PHC, JCLP, LMS, JMMC, NVRV, ESA, TGFMSL, CCB, RSS, LSCL, ASP, FRA, GMM, CMMV, KDC, RAMV
Project administration: RAMV
Software: TOB, PHC
Results: TOB, EBOJ, IJSF, JCLP, ESA, TGFMSL, CCB, RSS, LSCL, ASP, FRA, GMM, CMMV
Supervision: KDC, RAMV
Writing – original draft: TOB, EBOJ, IJSF, NVRV, FRA, RAMV
Writing – review and editing: TOB, EBOJ, IJSF, PHC, JCLP, LMS, JMMC, NVRV, ESA, TGFMSL, CCB, RSS, LSCL, ASP, FRA, GMM, CMMV, KDC, RAMV

Conflicts of Interest

None declared.

Multimedia Appendix 1
Representation of the number of procedures, distribution of priority care, and distribution of canceled requests.
[DOCX File, 210 KB - ojphi_v18i1e83494_app1.docx]

References

1. Del Nero CR. O que é economia da saúde. In: Economia Da Saúde [Book in Portuguese]: Editora Brasiliense; 1995:5-21 URL: https://repositorioslatinoamericanos.uchile.cl/handle/2250/9039046?show=full [accessed 2026-05-16]
2. Constituição da República Federativa do Brasil de 1988 [Article in Portuguese]. Presidência da República do Brasil. URL: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm [accessed 2026-05-16]
3. Barbosa DVS, Barbosa NB, Najberg E. Regulação em Saúde: desafios à governança do SUS [Article in Portuguese]. Cad saúde colet 2016;24(1):49-54. [doi: 10.1590/1414-462X201600010106]
4. Silva JPT. Análise da eficiência de hospitais regionais em um estado do Nordeste [Article in Portuguese]. Saúde debate 2019;43(120):84-97. [doi: 10.1590/0103-1104201912006]
5. Silva SN, Mello NFD, Ribeiro LR, Silva RED, Cota D. Implementation of health technologies in Brazil: analysis of federal guidelines for the public health system [Article in Portuguese]. Ciênc saúde coletiva 2024(29):e00322023. [doi: 10.1590/1413-81232024291.00322023]
6. Owen L, Morgan A, Fischer A, Ellis S, Hoy A, Kelly MP. The cost-effectiveness of public health interventions. J Public Health (Oxf) 2012 Mar;34(1):37-45. [doi: 10.1093/pubmed/fdr075] [Medline: 21933796]
7. OECD reviews of health systems: Brazil 2021 [Report in Portuguese]. : OECD Publishing; 2021 URL: https://www.oecd.org/en/publications/oecd-reviews-of-health-systems-brazil-2021_146d0dea-en.html [accessed 2026-05-16]
8. Cerda AA, García LY, Rivera-Arroyo J, Riquelme A, Teixeira JP, Jakovljevic M. Comparison of the healthcare system of Chile and Brazil: strengths, inefficiencies, and expenditures. Cost Eff Resour Alloc 2022 Dec 16;20(1):71. [doi: 10.1186/s12962-022-00405-9] [Medline: 36527052]
9. Lei n° 15121, de 10 de abril de 2025 [Article in Portuguese]. Presidência da República do Brasil. URL: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/L15121.htm [accessed 2025-06-05]
10. Iyamu I, Xu AXT, Gómez-Ramírez O, et al. Defining digital public health and the role of digitization, digitalization, and digital transformation: scoping review. JMIR Public Health Surveill 2021 Nov 26;7(11):e30399. [doi: 10.2196/30399] [Medline: 34842555]
11. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119. [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
12. Romans L. Computed Tomography for Technologists: Exam Review: Lippincott Williams & Wilkins; 2018.

XSL•FO
RenderX

13. Santos DLD. Distribuição da tomografia computadorizada e do grau de utilização do tomógrafo computadorizado no SUS [Dissertation] [Article in Portuguese]. : Universidade Federal da Bahia (UFBA); 2013 URL: https://repositorio.ufba.br/handle/ri/13127 [accessed 2026-05-16]
14. Leite HJD, dos Santos DL, Rasella D, de Souza SAL. Capacidade de produção e grau de utilização de tomógrafo computadorizado no Sistema Único de Saúde [Article in Portuguese]. Cad Saúde Pública 2014 Jun;30(6):1293-1304. [doi: 10.1590/0102-311X00140713]
15. Dovales ACM, da Rosa LAR, Kesminiene A, Pearce MS, Veiga LHS. Patterns and trends of computed tomography usage in outpatients of the Brazilian public healthcare system, 2001-2011. J Radiol Prot 2016 Sep;36(3):547-560. [doi: 10.1088/0952-4746/36/3/547] [Medline: 27460769]
16. Viacava F, Oliveira RD, Carvalho CDC, Laguardia J, Bellido JG. SUS: supply, access to and use of health services over the last 30 years [Article in Portuguese, English]. Cien Saude Colet 2018 Jun;23(6):1751-1762. [doi: 10.1590/1413-81232018236.06022018] [Medline: 29972484]
17. Valentim RADM, Lima TS, Cortez LR, et al. The relevance a technology ecosystem in the Brazilian National Health Service's Covid-19 response: the case of Rio Grande do Norte, Brazil [Article in Portuguese, English]. Cien Saude Colet 2021 Jun;26(6):2035-2052. [doi: 10.1590/1413-81232021266.44122020] [Medline: 34231717]
18. Barreto TDO, Veras NVR, Cardoso PH, et al. Artificial intelligence applied to analyzes during the pandemic: COVID-19 beds occupancy in the state of Rio Grande do Norte, Brazil. Front Artif Intell 2023;6(6):1290022. [doi: 10.3389/frai.2023.1290022] [Medline: 38145230]
19. Barreto TDO, Farias FLDO, Veras NVR, et al. Artificial intelligence applied to bed regulation in Rio Grande do Norte: data analysis and application of machine learning on the "RegulaRN Leitos Gerais" platform. PLoS One 2024;19(12):e0315379. [doi: 10.1371/journal.pone.0315379] [Medline: 39775276]
20. Panorama do censo 2022 [Web page in Portuguese]. Instituto Brasileiro de Geografia e Estatística (IBGE). 2022. URL: https://censo2022.ibge.gov.br/panorama [accessed 2025-06-05]
21. Bezerra MV. Análise da utilização da escala quick sequential organ failure assessment para tomada de decisão na regulação de leitos de UTI [Dissertation] [Article in Portuguese]. : Universidade Federal do Rio Grande do Norte; 2023 URL: https://repositorio.ufrn.br/server/api/core/bitstreams/2eaca7f4-1ac0-4933-8eba-54329b993005/content [accessed 2026-05-16]
22. Barreto TDO. Inteligência artificial aplicada ao ecossistema de regulação do estado do rio grande do norte (regularn): análises baseadas em machine learning em leitos COVID-19 e leitos gerais [Master's thesis] [Article in Portuguese]. : Universidade Federal do Rio Grande do Norte; 2024 URL: https://bdtd.ibict.br/vufind/Record/UFRN_265e6b665a954ba885216acf2484f22b [accessed 2026-05-16]
23. Cadastro Nacional de Estabelecimentos de Saúde (CNES) [Website in Portuguese]. URL: https://cnes.datasus.gov.br/ [accessed 2026-05-16]
24. DATASUS–TABNET [Website in Portuguese]. 2025. URL: https://datasus.saude.gov.br/ [accessed 2026-05-19]
25. RegulaRN outpatient computed tomography data. Zenodo. URL: https://zenodo.org/records/15830664 [accessed 2026-05-16]
26. Resolution no. 674, of May 6, 2022: provides for the classification of research and the processing of research protocols in the CEP/Conep System [Report in Portuguese]. : National Health Council / Ministry of Health; 2022 URL: https://clinregs.niaid.nih.gov/sites/default/files/documents/brazil/ResNo674_GoogleTranslate.pdf [accessed 2026-05-07]
27. Santos RCCD. Financiamento da saúde pública no Brasil [Report in Portuguese]. : Fórum; 2016 URL: https://www.cesau.ce.gov.br/wp-content/uploads/sites/161/2025/07/Financiamento-da-Saude-Publica-no-Brasil.pdf [accessed 2026-05-19]
28. Lei nº 8.080, de 19 de setembro de 1990 [Article in Portuguese]. Presidência da República do Brasil. URL: https://www.planalto.gov.br/ccivil_03/leis/l8080.htm [accessed 2025-05-05]
29. Lei Complementar nº 141, de 13 de janeiro de 2012 [Article in Portuguese]. Presidência da República do Brasil. URL: https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp141.htm [accessed 2025-05-05]
30. Pucci de Sá e Benevides R, Funcia F. Desafios para melhorar a qualidade dos gastos do SUS. In: SUS: Avaliação Da Eficiência Do Gasto Público Em Saúde [Book in Portuguese]: Instituto de Pesquisa Econômica Aplicada (IPEA); 2023:282-315 URL: https://repositorio.ipea.gov.br/server/api/core/bitstreams/6afd2e86-3545-4ab7-a24e-02db9390eaef/content [accessed 2026-05-19]
31. Nunes Júnior NB, Pacheco MAB. The cost-effectiveness of imaging examinations as a parameter of efficiency in a health service in the diagnosis of community acquired pneumonia. Saúde Coletiva 2023(13-86):12537-12560. [doi: 10.36489/saudecoletiva.2023v13i86p12537-12560]
32. Marinho A, Funcia FR, Vianna Sobrinho L. SUS: Avaliação Da Eficiência Do Gasto Público Em Saúde [Book in Portuguese]: Instituto de Pesquisa Econômica Aplicada; 2023. URL: https://repositorio.ipea.gov.br/entities/book/7c63b81e-2306-4db1-b2ad-791a6ea529da [accessed 2026-05-19]
33. Dang W, Kielar AZ, Fu AYN, Chong ST, McInnes MD. Does distance matter? Effect of having a dedicated CT scanner in the emergency department on completion of CT imaging and final patient disposition times. J Am Coll Radiol 2015 Mar;12(3):277-283. [doi: 10.1016/j.jacr.2014.11.003] [Medline: 25640574]
34. de Carvalho EM, Costa e Silva FV, dos Santos PR. Planning and administration of environmental security and care in health services [Article in Portuguese]. R pesq cuid fundam online 2018;10(1):224-232. [doi: 10.9789/2175-5361.2018.v10i1.224-232]

XSL•FO
RenderX

35. Bittencourt O, Verter V, Yalovsky M. Hospital capacity management based on the queueing theory. Int J Product Perform Manag 2018 Feb 12;67(2):224-238. [doi: 10.1108/IJPPM-12-2015-0193]
36. de Almeida Botega L, Andrade MV, Guedes GR. Brazilian hospitals' performance: an assessment of the Unified Health System (SUS). Health Care Manag Sci 2020 Sep;23(3):443-452. [doi: 10.1007/s10729-020-09505-5] [Medline: 32372264]
37. Chandra D, Tiwari A. A systematic survey on different modeling methods to manage waiting time in ambulatory patient services. Int J Eng Appl Sci Technol ;4(1):14-23. [doi: 10.33564/IJEAST.2019.v04i01.003]
38. Al-Ashi R, Awad M, Shamayleh A, Alzaatreh A. Medical equipment maintenance performance and influential factors. J Qual Maint Eng 2026 Mar 10;32(1):180-204. [doi: 10.1108/JQME-04-2024-0042]
39. Nasir UM, Tijjani AM, Mohammed G, et al. Quantifying radiation dose and risk in diagnostic X-ray and computed tomography imaging procedures. Gombe J Adm Manag 2025;8(4B):63-75. [doi: 10.56892/bima.v8i4B.1169]
40. Metodologia de administração de dados [Report in Portuguese]. : Departamento de Informação e Informática do SUS (DATASUS); 2023 URL: https://datasus.saude.gov.br/wp-content/uploads/2023/12/Metodologia-de-Administracao-de-Dados-MAD.pdf [accessed 2025-06-28]

Abbreviations

ADCT: Act of Transitory Constitutional Provisions
APAC: Outpatient Health Procedure Authorization
CT: computer tomography
DATASUS: Department of Informatics of the Unified Health System
PET: positron emission tomography
SIGUS: Outpatient Regulation System
SISREG: National Regulation System
SUS: Brazilian National Health System

Edited by E Mensah; submitted 08.Sep.2025; peer-reviewed by F Zicker, L Shrestha; revised version received 06.Mar.2026; accepted 04.May.2026; published 01.Jun.2026.

Please cite as:
Barreto TDO, de Oliveira Júnior EB, dos Santos Felipe IJ, de Lima Paiva JC, Rodrigues Veras NV, de Macedo Silva L, Macedo da Costa JM, Santos de Aragão É, Góis Farias de Moura Santos Lima T, da Costa Barros C, Silveira e Silva R, Rocha de Andrade F, Holanda Cardoso P, Santos Pinheiro A, Medeiros Machado G, Miranda Veloso CM, Simões Camara Leão L, Dantas Coutinho K, de Medeiros Valentim RA
The Use of Tomographs in Brazil's National Health System: Case Study on the Efficiency of the Public Network in Rio Grande do Norte
Online J Public Health Inform 2026;18:e83494
URL: https://ojphi.jmir.org/2026/1/e83494
doi:10.2196/83494

© Tiago de Oliveira Barreto, Elinaldo Bernardo de Oliveira Júnior, Israel José dos Santos Felipe, Jordana Crislayne de Lima Paiva, Nicolas Vinicius Rodrigues Veras, Lorena de Macedo Silva, João Maria Macedo da Costa, Érika Santos de Aragão, Thaísa Góis Farias de Moura Santos Lima, Célio da Costa Barro, Rafael Silveira e Silva, Fernando Rocha de Andrade, Pablo Holanda Cardoso, Andréa Santos Pinheiro, Guilherme Medeiros Machado, Claudia Maria Miranda Veloso, Luciana Simões Camara Leão, Karilany Dantas Coutinho, Ricardo Alexsandro de Medeiros Valentim. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 1.Jun.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Workflow-Based Information Management Framework for Multicenter Research Studies: Design and Development

Hasan Sulaeman[1], MSci; Mars Stone[1], PhD; Roberta Bruhn[1], PhD; Karla Zurita[1], BA; Anh Nguyen[1], MSc; Vincent Chiang[1], BS; Jefferson Michael Jones[2], MD, PhD; Xutao Deng[1], PhD; Brian Custer[1], MPH, PhD; Michael Busch[1], MD, PhD; Eduard Grebe[1], PhD

[1]Vitalant Research Institute, 360 Spear Street, San Francisco, CA, United States

[2]Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases, Atlanta, GA, United States

Corresponding Author:
Hasan Sulaeman, MSci
Vitalant Research Institute, 360 Spear Street, San Francisco, CA, United States

Abstract

Background: Biological and health research is increasingly data-driven, with commercial and academic institutions generating data at unprecedented rates. The rapid pace of data generation, together with lessons learned during the COVID-19 pandemic, underscores the need for nimble, transparent, and dependable data infrastructures that enable rapid study execution and timely insights to inform public health policy and practice.

Objective: This paper describes the workflow-based information management (WIM) framework, a flexible research information management system designed to support diverse epidemiologic workflows and data-intensive research projects.

Methods: WIM was developed as a modular, workflow-oriented framework built on the open-source R (R Foundation) programming language and its extensive ecosystem of community-developed packages. The framework emphasizes reproducibility, adaptability, and transparency, enabling users to design and manage research workflows tailored to specific study requirements. We describe the architecture and core components of WIM and illustrate its application through representative epidemiologic research scenarios.

Results: The framework supported high-volume, multiorganizational research; managing >3.7 million donation and testing records from 17 blood collection organizations across the United States. The WIM framework was readily adapted to a wide range of epidemiologic studies and research projects, demonstrating flexibility across varying data types, analytical needs, and operational contexts. By leveraging established R-based tools and workflows, WIM supported efficient data ingestion, processing, analysis, and reporting while promoting reproducible and collaborative research practices. The framework facilitated rapid iteration and reuse of workflows, addressing common challenges in managing complex and evolving research studies.

Conclusions: WIM provides a flexible, open-source, and extensible approach to research information management for modern biological and health research. By integrating workflow-based design principles with the R ecosystem, the framework supports reproducible analysis, scalable research operations, and rapid study execution. WIM offers a practical solution for institutions seeking adaptable data infrastructure to support epidemiologic research and inform public health decision-making.

(Online J Public Health Inform 2026;18:e81119) doi:10.2196/81119

KEYWORDS

multicenter studies; large studies; open source; data system; data management; data infrastructure; ETL; extract transform load

Introduction

With technological advances in recent decades, bioinformatics and data management are becoming increasingly important to life sciences research [1-3]. Contemporary biological research is often dependent on the management, sharing, and analysis of large-scale, aggregated data, in particular for studies designed to tackle large scientific and societal issues such as the COVID-19 pandemic [4-6].

During the COVID-19 pandemic, scientific interest in specific public health questions shifted as the pandemic progressed [7]. For example, early seroprevalence studies were performed to determine the proportion of the population that had been infected with SARS-CoV-2. After the introduction of vaccines, these had to be modified to determine the proportion of the population that had been vaccinated, infected, or both [8-10]. In May 2020, the US Centers for Disease Control and Prevention (CDC), in partnership with Vitalant Research Institute (VRI), established the Nationwide Blood Donor Seroprevalence study (NBDS). In July 2020, NBDS launched the first phase of the study in

XSL•FO
RenderX

collaboration with 17 blood collection organizations across the United States and Puerto Rico. Multiple testing laboratories, including VRI and Creative Testing Solutions, captured, tested, and analyzed approximately 150,000 blood donation specimens monthly in a serial cross-sectional seroprevalence survey. As the pandemic evolved, a large proportion of the population had infection-induced antibodies or had been vaccinated. To determine if infections occurred before or after vaccination or to detect reinfection, longitudinal data is required. The need to detect reinfections and infections in vaccinated individuals led to the launch at the start of 2022 of the second phase of the program: the Nationwide Blood Donor Cohort study (NBDC). The NBDC program switched from a cross-sectional to a longitudinal study format to follow a cohort of blood donors from BCOs to address questions such as the incidence of infection (in vaccinated and unvaccinated individuals) and multiple sequential infections (reinfections) with SARS-CoV-2, waning antibody titers following vaccination or infections, and correlates of protection against SARS-CoV-2 infection [11,12]. Given the fluid nature of these large studies, a secure and nimble data infrastructure was needed to meet the studies' needs, including management of large quantities of data on donors and donations from BCO records, serologic testing data, and responses to electronic donor surveys.

BCOs are primed for nationwide studies such as the NBDS and NBDC, with a physical infrastructure already in place to collect and test biospecimens from blood donations, including capturing residual specimens after routine blood screening for additional research testing and executing electronic surveys of participating donors. However, the data management framework and infrastructure to facilitate research programs built on the blood collection system would benefit from further development, a challenge many other organizations face and are attempting to tackle [13,14]. Due to the volume and complexity of the NBDS and NBDC data, traditional methods of data manipulation and management using software such as spreadsheet-based software were not an option. Though there are numerous commercially available data and project management software systems, their use would require a costly upfront license fee and/or a monthly cost incurred from services rendered for a software-as-a-service platform. Furthermore, off-the-shelf solutions typically require extensive customization to meet the needs of large and complex research programs.

Fortunato and Galassi (2021) [15] defined open-source software as "any computer program released under a license that grants users rights to run the program for any purpose, to study it, to modify it, and to redistribute it in original or modified form." R and Python (Python Software Foundation) are 2 popular data manipulation and analysis programming languages that fall into the category of open-source software. We chose to build our data management infrastructure in R due to its popularity among life science professionals, versatility, low barrier to entry, and a strong community of developers that has built an ecosystem of data-related packages [16,17]. Using the same principles as the framework we detail here, an analogous data system could easily be built in Python or any other open-source general programming language.

The design of our workflow-based information management (hereafter WIM) framework follows a few distinct design principles to achieve the goal of a nimble and reusable informatics management system that could reliably handle large amounts of data using open-source programming languages. First, our framework follows the FAIR (Findable, Accessible, Interoperable, Reusable) principles detailed in Wilkinson et al (2016) [18] for data reusability, accessibility, and system interoperability. Second, our framework follows the design principle of loose coupling, wherein components are only weakly associated with each other or the system, and so changes in 1 component least affect the performance of other components [19]. This principle allows the framework to be agnostic regarding operating systems and underlying platforms such as the relational database management system. Lastly, due to the fluidity of research studies, to prioritize reusability and adaptability of the framework, we opted not to have a formal component model [20].

Methods

Concept

The framework is built on modules, some automated, used in the same way functions from packages are used in R, and is publicly accessible on GitHub (GitHub, Inc) [21]. The modules are called and managed using the *box* package, which negates the need to import each function or to publish the suite of modules as an R package [22]. Additionally, all settings used by the modules are set using a single project-wide configuration file in the yet another markup language (YAML) format, which includes information on quality control (QC) processes, data dictionaries used by the study, credentials where required to run certain processes, database connection strings, and any other pertinent information used in data ingestion and reporting [23]. The use of a configuration file streamlines the implementation of changes in data flow, formatting, and reporting and lessens the time required to apply any changes to the data system, while also allowing multiple data managers to execute the same functions and scripts with any individual configurations managed in a separate configuration file.

The modules are platform-agnostic and can be run on either a cloud platform or locally with little to no input, depending on how much of the process requires manual review. They could also be run on a set schedule as a *cron job*, a system for scheduling specific tasks on Unix-based operating systems (eg, Linux [DragonByte Technologies] and macOS [Apple Inc]), or automatically when required using a listener to respond to an external action. In our case, for pragmatic reasons, the modules are run locally on desktop machines by one or more administrators. Separately, though modules run processes through R, an external connection is required to run certain processes, such as performing data retrieval from an external source, monitoring a secure file transfer protocol (sFTP) server for new data deliveries, writing processed and quality-controlled data to a central database, and transferring data to other organizations. These connections are made in R using certain packages, including *curl*, *RCurl*, and *sFTP* [24-26].

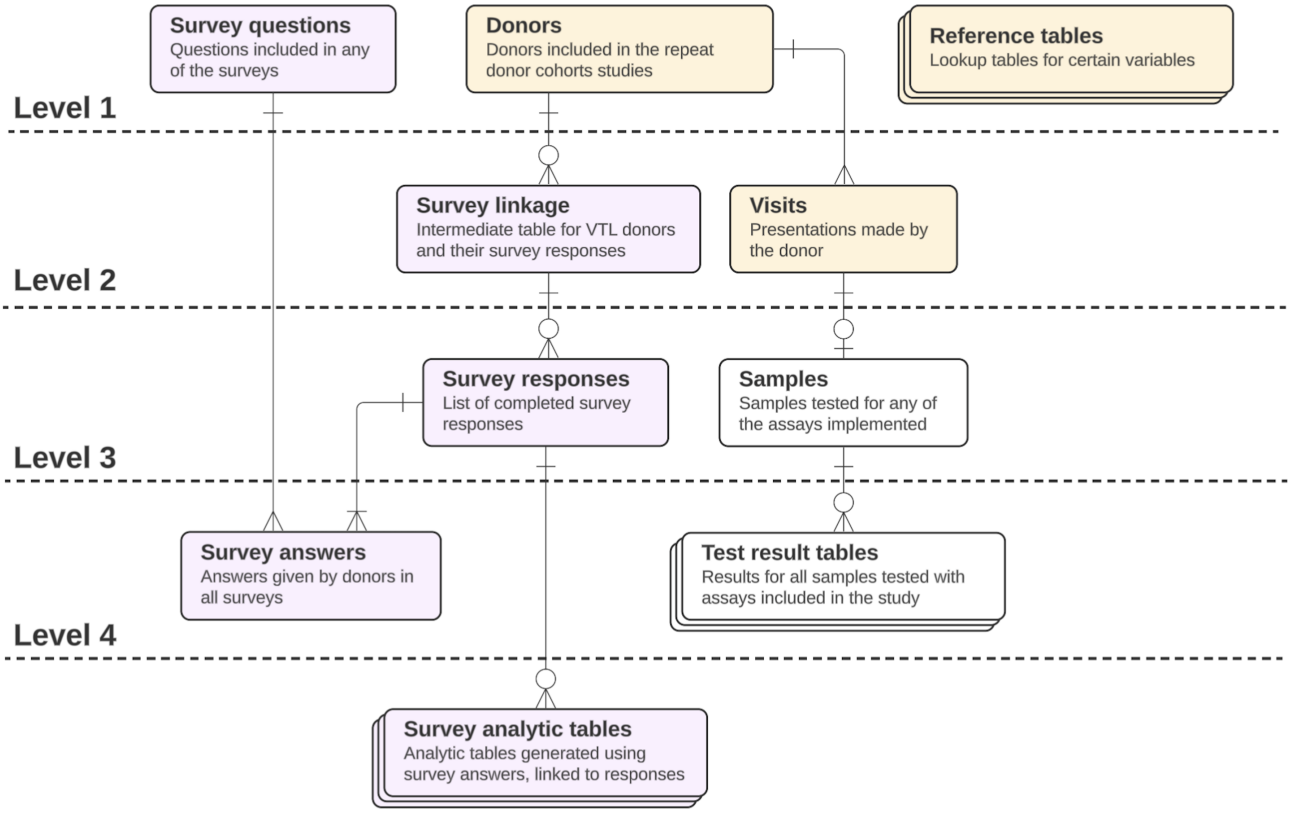
XSL•FO
RenderX

Lastly, a central goal in the implementation is for all relevant stakeholders to have access to both the managed central database and the version-controlled code repository, managed using Git. To lower the technical skill floor required for accessing this study's database, a custom front-end was built using Python and Flask (Pallets), which caters specifically to the organization and stakeholders' data needs and is abstracted. The Git repository for the modules includes all the information and scripts required to install and run the modules, with users having only to install the required dependencies using a script included in the Git repository.

Data Sources and Types

Data sources for the NBDS and NBDC included donor, donation, and blood specimen testing data (Figure 1). When the NBDC study launched, donor electronic survey data were added as a data source [27,28]. Donor and donation data for the program are extracted from BCO operational data systems. Donor data includes time-invariant donor characteristics such as birth date and blood group, while donation data includes information collected at each donation event, for example, the blood collection procedure, responses to routine questionnaires administered at the time of donation, as well as other data points that might differ for a given donor over time. Data points such as ZIP code of residence, sex, race, and ethnicity were reported by the donor at the time of donation. It is worth noting that reference tables, also known as lookup tables, were generated and used for tracking changes to donor identifiers and certain donation-level data where the original data point from the source should be kept, along with any study-specific interpretation we applied. For example, the way organizations record and group donor race and ethnicity data might differ from how a study groups these data. By implementing a lookup table that links the original data to the study-specific data grouping, stakeholders can revise groupings, if necessary. Testing data includes all valid serological test results from assays in the program, whether for routine study-directed testing or for assay validation substudies [29,30]. Specifications for which variables are captured, stored, and reported differ between assays and are dependent on the assay manufacturer's instructions for use and determinations of the most relevant measurements.

Figure 1. Simplified ERD for the repeat donor cohorts program. Levels are used to imply dependencies between the different tables. Purple, yellow, and white are used to denote survey data, donation data, and testing data, respectively. Standard cardinality notation is used. ERD: entity relationship diagram.



Lastly, donor survey data for the NBDC can be split into 2 other data types: survey question information and survey response information—both of which are equally important to properly manage. Survey questions for the program often change between different survey rounds to accommodate the changing needs of this study, so there is a need to manage question information while tracking question equivalency across survey rounds. Due to survey questions sometimes changing, survey responses, defined as a completed survey form, must be managed separately from the discrete answers to a question. This is so that answers to equivalent questions across survey rounds can be parsed and analyzed without requiring excessive processing time or manual recoding during analysis.

Data Dictionary

The NBDS program accepted frequent data submissions from 17 BCOs and multiple testing laboratories across the United States and Puerto Rico. The NBDC program, on the other hand,

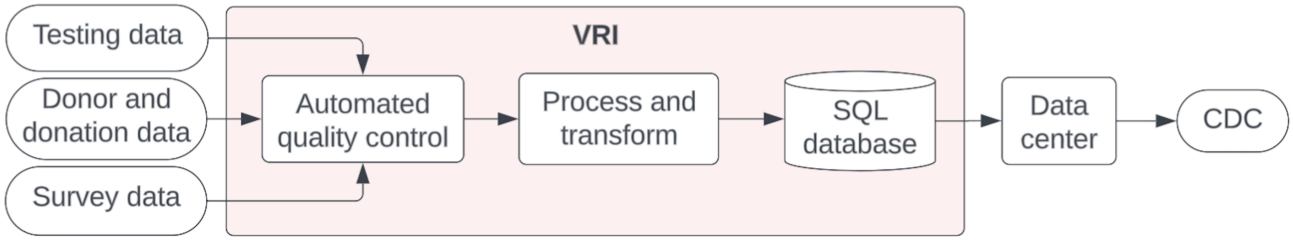
XSL•FO
RenderX

while only accepting submissions from 2 organizations and 2 testing laboratories, expanded the scope and quantity of data collected, stored, and managed for its cohorts. Both studies required comprehensive data dictionaries, regularly updated to reflect changes made to study procedures and methods, and detailing how to format data for each type of data transfer between organizations, including what values or valid ranges are acceptable for each field. Data dictionaries for both the NBDS and NBDC programs are included in Multimedia Appendices 1 and 2.

Data Flow

Though the data flow for the NBDS and NBDC programs differed slightly due to different data sources, types, and testing algorithms applied to biospecimens, the general flow of data remained the same (Figure 2). Data from all sources first flowed to VRI, where it underwent QC and, if the submission passed quality control procedures, was imported to this study's database. The data submission was then reported to the data coordinating center (DCC), the contract research organization Westat, where it underwent a second QC step, where primary analysis took place, and was finally transferred to the sponsor (CDC).

Figure 2. Workflows defined by the framework. Workflows include an extract, transform, load process from each data source, a process to manage data while at rest, and a process for reporting to the data center and later the US Centers for Disease Control and Prevention. CDC: US Centers for Disease Control and Prevention; VRI: Vitalant Research Institute.



For the NBDS, to accommodate varying testing and sample flows, participating BCOs were binned into groups, each of which had a distinct data transmission format and schedule [12]. Groups evolved with the program and adapted to changes in the testing algorithm and the capacity of testing laboratories. For the NBDC, data flow differed by data source and organization. Donor and donation data for Vitalant flowed to VRI, while for ARC, the submission went directly to the data broker. Testing results are always routed through VRI before going to the DCC, while survey results were reported directly to the DCC for both Vitalant and ARC (Figure 2)

Study-Specific Relational Database

The use of a study-specific database was critical in the day-to-day operations of both the framework and this study itself (Figure 2). Having certain restrictions inherent in the structure of the relational database, which included length of the field, primary and foreign keys, data types, allowable potentially identifying information, and allowable codes for categorical fields (enumerated types), establishes this study's database both as the canonical source of study data and as a redundant QC process by passively ensuring data ingested is in accordance with this study's data dictionary as well as all relevant regulations surrounding human-participants research and privacy protections. In our case, with the hierarchical structure of donor, visit, testing, and survey data, a study database was instrumental in keeping data integrity across tables. Lastly, it is important to note that for organizations working with potentially identifying information, such as hospital records, granular data governance is crucial to augment privacy protections by allowing only relevant stakeholders to view certain parts of the database.

Automated Secure Data Retrieval

This study relied on 2 data transfer methods that are encrypted both in transit and at rest: Microsoft OneDrive (Microsoft Corp) and an sFTP server. The module responsible for automated secure data retrieval performs several processes and uses Microsoft Workflows (previously Microsoft Power Automate; Microsoft Corp) as a method of notifying submitters and stakeholders of any new submissions. The workflow or module of the framework is also responsible for notifying stakeholders whether the submission passed QC and was accepted, or did not pass the QC checks and was rejected. Each participating organization is given a separate directory (also known as a bucket) on the sFTP server, along with credentials for their account. Within each bucket, 3 directories are made: upload, download, and archive. Users are only allowed to add or remove files in the upload and download directories, where they can upload data submissions or download data validation reports and rejected data submissions, respectively. The archive folder is where accepted submissions are moved and stored as a backup and for data audit purposes.

Automated listeners monitor each directory, with a process that is triggered when a file is altered or added in the directory. A change in the upload directory triggers a new submission email to all relevant stakeholders, while a change in the download and archive directories triggers rejection and acceptance emails, respectively. When a new submission is added to the upload folder, an R module downloads the file and removes it from the upload directory. The module then performs a QC process specified in the configuration file and, depending on the outcome of the quality control check, performs 1 of 2 actions: if the submission failed QC, the module uploads an itemized list of all quality control issues encountered to the download folder, which prompts a submission rejection email to all stakeholders with the itemized list of quality control issues encountered attached. Data type, length, date, and enumerator checks are examples of what the quality control process entails. If the submission passes quality control, the accepted submission is uploaded to the archive folder, the data is imported into this

XSL•FO
RenderX

study's database, and a submission acceptance email is sent out to all stakeholders.

Data Intake Workflows

Donor and Donation Data

The extract, transform, load (ETL) process for donor and donation data differed between the NBDS and NBDC programs, as well as by organization for both studies. For the NBDS, donation data was transferred from all participating organizations through the sFTP server, while for the NBDC, the data for Vitalant was transferred through Samba servers to VRI. Donor and donation data for ARC were sent directly to the DCC. For the NBDS, all submissions were required to adhere to the agreed-upon data dictionary and were submitted in accordance with a data submission guideline document shared with all participating organizations (Multimedia Appendix 3). For the NBDC, the only processes required were QC and data transformation for Vitalant donor and donation data. Once Vitalant's data submission passed quality control and was transformed per this study's data dictionary, the data were imported to the study database (Multimedia Appendix 2).

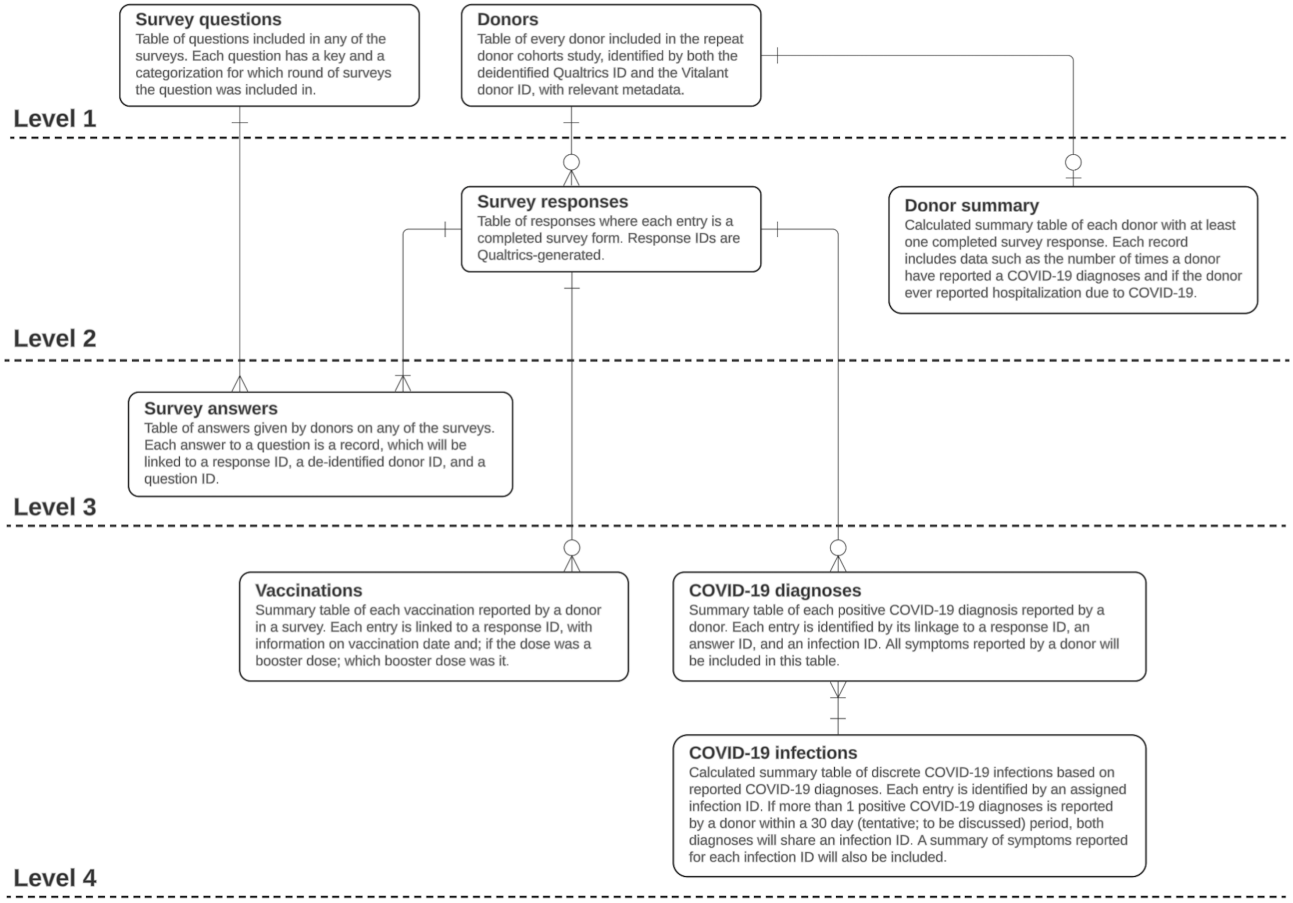
Laboratory Testing Data

Testing data for both the NBDS and NBDC programs were generated by testing and immunology laboratories, with each testing laboratory having a separate workflow and R module. Transmission of test results from the test laboratory was sent via a cloud service, and a QC check was performed to ensure adherence to this study's data dictionary. For testing data from the immunology laboratory, output from testing instruments was retrieved, reformatted to fit this study's data dictionary specifications, and imported into this study's database automatically by an R module developed for this purpose (Figure 2).

Donor Survey Data

Donor survey data consisted of both survey questions and responses over time. As survey questions change to accommodate changes to this study over time, the data management approach must track how these questions shift and change over time—maintaining equivalency of questions across survey rounds. The data management approach also must be able to record answers to each question without having to change the database table structure with every change in survey question (Figure 3). It was also important to capture analysis-ready data that integrated data captured across survey rounds in a consistent format.

Figure 3. Management of survey data in this study's database. Levels are used to imply the linkage dependency of tables. Level 1 tables do not need primary and foreign keys to link to another table, while levels 2, 3, and 4 rely on linkage with subsequent levels. Standard cardinality notation is used.



The database structure for the donor survey data was generally categorized as either a raw data table or an analytic table. The raw data tables consist of survey questions, survey responses, and survey answers, while analytic tables include any tables using the survey data to derive information that might be beneficial for analysts without having to parse the raw answers.

XSL•FO
RenderX

The survey questions table records every question ever given to donors across all survey rounds and assigns to each question a question ID. Equivalent questions across rounds share the same question ID for ease of analysis. The survey responses table records whether a respondent has a completed response for each survey round and whether the respondent consented for their answers to be used in research (ie, 1 row per respondent per survey round). For donors who did not consent for their answers to be used in research, none of their survey answers are recorded in the database, though they are still recorded as having responded to the survey and flagged as nonconsenting. While respondents generally did not answer any further questions if they failed to provide consent at the outset, consent could be withdrawn later in the process, and these technical safeguards ensured that no responses were recorded without consent to use the responses in research. The survey answers table consists of several fields, including a unique response ID that links to the responses table, a question ID that links to the survey questions table, and the answer given to the specific question by the respondent on the survey.

Two modules are responsible for managing the survey data, both of which rely on raw output from the survey platform. The first module is responsible for assigning equivalency to each question based on the question ID and for updating the survey questions table, while the other module parses survey responses and updates both the survey responses and survey answers table.

Reporting Workflows

Reporting is done through a module for each type of report defined by the program. For example, reporting for the NBDS program only required testing reports. In contrast, the NBDC program required donor, donation, survey, and testing reports. While each reporting module pulls information from the configuration file on which fields to extract from which tables in the database, and how to transform the data to fit the data dictionary's reporting specifications, extra code is generally needed for specific requirements that require more specific QC steps than the standard checks for data format and length. This necessitates more modular code to be included. For example, changes in how a laboratory test is configured might require changes in the quality control process, for example, to check for values in other fields, or to transform values based on one or more entries in other fields.

Ethical Considerations

All blood donors consented to the use of deidentified, residual specimens for further research purposes. Consistent with the policies and guidance of the University of California, San Francisco Institutional Review Board, VRI self-certified the use of deidentified donations in this study as not meeting the criteria for human participant research. CDC investigators reviewed and relied on this determination as consistent with applicable federal law and CDC policy (45 C.F.R. part 46, 21 C.F.R. part 56; 42 U.S.C. § 241[d]; 5 U.S.C. § 552a; 44 U.S.C. § 3501). The study number is Pro00056783. The donor surveys conducted by Vitalant were conducted under a protocol supervised and approved by the Advarra Institutional Review Board (Pro00056783) and linked to biospecimens in deidentified form.

Results

Customizability

The framework we developed resulted in a data system that is robust, reusable, and adaptable. Several factors contribute to the adaptability and reusability of the system. First, by separating generic code from the study-specific information, the system can adapt to changes in the study or be reused for another study with a greatly reduced investment of time to set up a study-specific data system (Table 1). The use of a configuration file that instructs the base code on how to access certain study-specific resources; what data types, variable lengths, and data types are allowed for specific entries; and how to generate multiple types of reporting required by the study or project allows for the study's data system to be quickly adapted to changes in the study or implementation of new substudies, without lengthy development time and making extensive changes to the code. An example would be how changes to a study's codebook are implemented. For systems where certain fields or information are hard-coded, implementing a change means changing the base code. In the case of our system, editing the configuration file is sufficient to implement the changes.

XSL•FO
RenderX

Table . Overview of the key design principles that guided the development of the WIM[a] data management framework.

Principle	What the principle seeks	WIM design features	Examples from NBDS[b]/NBDC[c]
Findable	Data and metadata can be easily found via clear documentation and identifiers	Comprehensive data dictionaries and persistent identifiers are enforced across different studies	Shared IDs kept consistent across NBDS and NBDC to enable cross-study linkage and downstream analyses
Accessible	Users can retrieve data with appropriate authorization and governance	Abstracted database web frontend with authentication and granular access control, lowering the skill floor while supporting governance	Nontechnical collaborators access curated tables and reports through the web frontend, with role-based permissions
Interoperable	Systems, tools, and data can work across platforms and contexts	Platform-agnostic modules can connect to multiple software platforms; deployable on-prem or cloud	ETL[d]/QC[e] modules run on Linux or Windows (Microsoft Corp) and interface with different database engines and file exchange endpoints,
Reusable	Data and tools are packaged to maximize reuse across studies and time	Single project-wide YAML[f] configuration externalizes study-specific logic (ie, dictionaries and QC rules) from core code. Omission of a formal component model to preserve flexibility	Reused ingestion, QC, and reporting modules between NBDS and NBDC with primarily configuration changes, minimizing code edits and reducing setup time
Loose coupling	Components change with minimal ripple effects on others	Strict modularization (ingestion, QC, ETL, and reporting separated). Configuration- driven behavior	Updating a QC rule or adding a new data source did not require edits to ingestion listeners or reporting modules

[a]WIM: workflow-based information management.

[b]NBDS: Nationwide Blood Donor Seroprevalence.

[c]NBDC: Nationwide Blood Donor Cohort.

[d]ETL: extract, transform, load.

[e]QC: quality control.

[f]YAML: yet another markup language.

Second, for data with significant structural variability over time, such as the survey data in our case, native functionality of database engines should be leveraged (Figure 3). For example, survey data for our study changed regularly between survey rounds, and keeping the data in wide format would mean having to manage a very wide table. As a wide table would mean having to change table configurations and column names as changes are implemented to the surveys, we elected to store survey data in a long format—with discrete tables for survey questions, responses, and answers. The use of long format for both survey answers and questions negated the need for tables with a column for each question ever asked on the survey. Additionally, having both the survey questions and answers in a long-format table means that the general data structure can be preserved, even if changes are made to the survey questions, and that the survey questions table can be used as a reference table to check for equivalency of specific questions between different rounds of the survey.

The last factor to contribute to the framework's reusability and adaptability is being flexible with respect to connections with other systems, along with loose coupling. Interoperability, as defined by Wilkinson et al (2016) [18], is the ability of data or tools from noncooperating resources to integrate or work together with minimal effort. By using and interoperating with other preexisting data systems and platforms, including proprietary software systems, the framework can fill in gaps in functionality and be implemented in parallel with both legacy and new systems.

Evaluation of Use Case

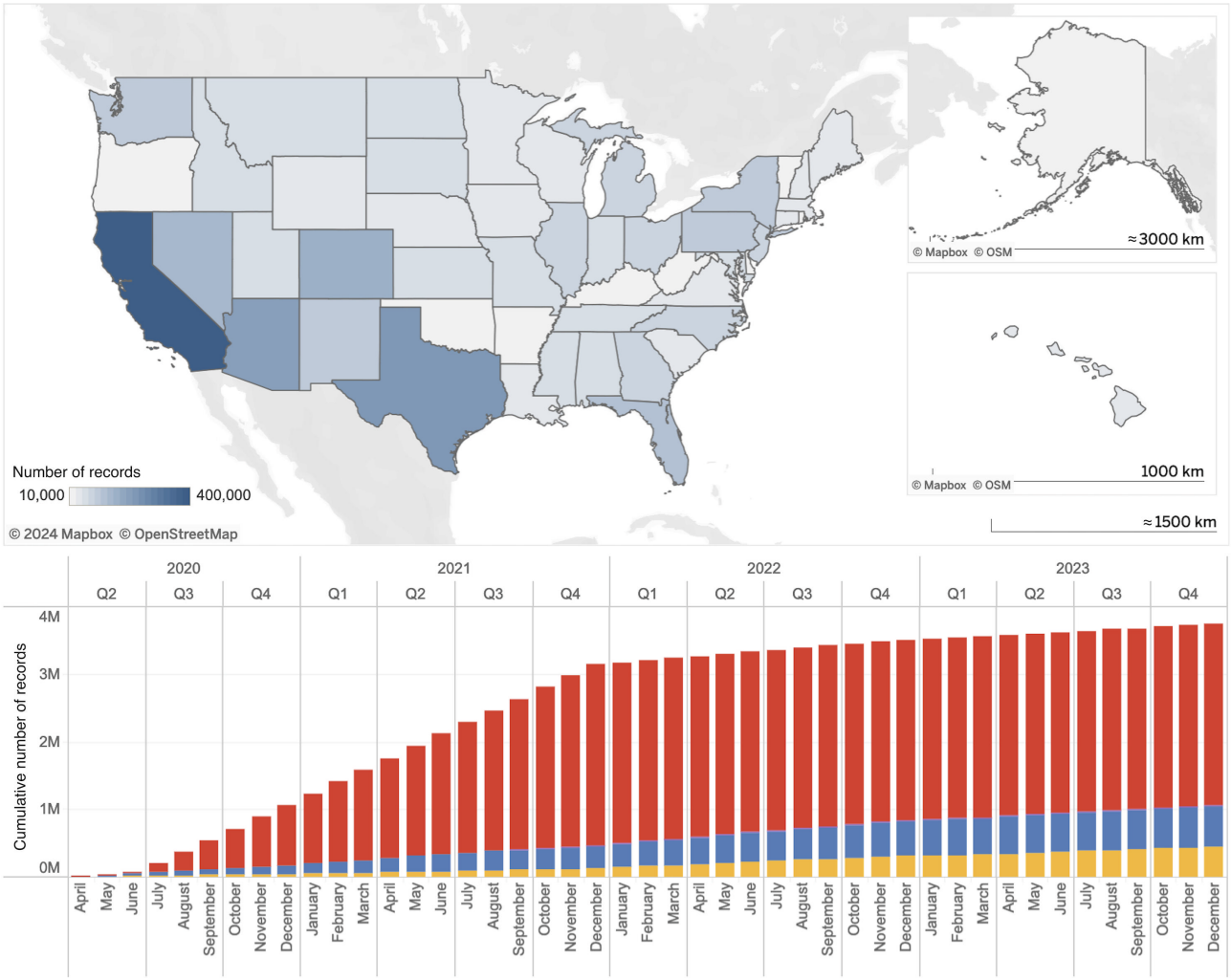
We used deployment efficiency, adaptability, and update latency, data throughput, and reusability of core modules as key operational indicators used to evaluate the framework. These metrics reflect real-world performance under high-volume, multiorganizational research conditions. Taken together, these metrics highlight WIM's ability to support large, complex, and rapidly evolving research programs by providing a scalable, adaptable, and reusable data-management infrastructure.

Data Throughput and QC Performance

For the NBDS study, our study data infrastructure handled 2,670,225 donation and testing records from all 50 states in the United States over the lifetime of this study (Figure 4). These records were generated by 17 BCOs and were submitted to VRI monthly. For the NBDC study, our study data infrastructure handled a total of 1,064,381 donation records from 65,524 donors who were longitudinally followed over the course of 2 or more years (Figure 4). Donation and donor data were generated and updated on a quarterly basis, while testing data were updated on a weekly basis.

XSL•FO
RenderX

Figure 4. Number of records collected and managed by the NBDS and NBDC programs. (A) Number of records collected by state, with 270,319 records excluded due to missing geographic information. (B) Number of records accumulated over the two programs' lifetime. Colors yellow, blue, and red denote whether the record was collected for the repeat donor cohorts 2023, Vitalant repeat donor cohorts 2022, or the National Blood Donor Serosurveillance study, respectively. NBDC: Nationwide Blood Donor Cohort; NBDS: Nationwide Blood Donor Seroprevalence.



Deployment Efficiency

For deployment efficiency, initial system deployment required approximately 6 months for the NBDS program, whereas deployment for the more complex NBDC program required just over 2 months. This reduction in setup time reflects the reusability of core modules and the ability to externalize all study-specific logic in the YAML configuration file, allowing new implementations to leverage an existing codebase with minimal modification. For other studies with different data needs, only study-specific logic and QC checks would have to be developed or adapted before implementation.

Adaptability and Update Latency

Operational and methodological changes to this study, such as updates to data dictionaries, QC criteria, reporting formats, or survey structures, can be implemented typically within 1 day for minor revisions and within a few days for more complex changes involving changes to multiple workflows or a drastic overhaul of study design. As WIM isolates study-specific rules from core modules and uses an external configuration file, updates could be integrated rapidly into ETL, QC, or reporting pipelines with little to no change to the underlying software.

The framework successfully supported very high data volumes, which include 2.67 million donation and testing records in NBDS and over 1 million longitudinal donation records in NBDC. Data ingestion and reporting vary in size and complexity, but the framework handled data ingestion and reporting tasks of up to 500,000 rows seamlessly. Automated QC workflows processed each data submission immediately upon arrival, applying rule-based checks for field formats, enumerated values, date validity, and cross-table integrity. The use of automated listeners and rule-based validation eliminated the need for manual prescreening and ensured rapid feedback to submitting organizations.

Reusability of Core Modules

For the NBDS and NBDC studies highlighted here, processes such as ETL routines, QC, secure data retrieval, and reporting were reused with no modification. Study-specific customization was largely restricted to configuration files, data dictionaries, and survey-specific logic. This high degree of functional reuse demonstrates the portability of the framework and supports efficient deployment in new research settings.

XSL•FO
RenderX

Discussion

Overview

Modern data generation techniques in life science research increasingly result in large quantities of data. With the increasing rate of data generation comes an increasing need for reliable, stable, and sophisticated information management systems to manage, store, and share data [2,31,32]. When large datasets from multiple, often disjointed, data sources can be accommodated, our ability to efficiently study large societal and scientific challenges is greatly increased (eg, studying population-wide effects of the COVID-19 pandemic) [33-35]. In this paper, we describe how our approach of developing a modular, flexible, and interoperable data infrastructure using open-source tools allowed us to nimbly manage data in rapidly evolving nationwide COVID-19 studies during the pandemic, and how we can scale, reuse, and adapt the same framework to manage other large multicenter studies with minimal changes to the core software.

Despite the availability of both commercial and open-source data-management platforms, the WIM framework provides several contributions that differentiate it from existing systems and were essential to supporting the rapidly evolving NBDS and NBDC research programs. First, it uses a single project-wide YAML configuration file to control all modules responsible for data ingestion, transformation, QC, reporting, and database interactions. This design places study-specific logic entirely outside the core codebase, allowing substantial changes to data dictionaries, quality control rules, or reporting specifications to be implemented without modifying the underlying software. This approach reduces development time, lowers operational complexity, and enables modules to be reused with minimal reconfiguration across multiple studies. Second, WIM achieves a high level of modularization and loose coupling, in which each workflow module operates independently. This stands in contrast to many proprietary systems that require vendor-managed customization or open-source pipelines that integrate ingestion, processing, and reporting in a rigid or monolithic structure. By separating generic functions from study-specific rules, WIM supports rapid iteration as research needs evolve and minimizes unintended downstream effects when updating a single module.

WIM's architecture is designed for portability and reusability beyond the described COVID-19 serosurveillance programs described. Core modules for secure data retrieval, ETL processes, QC, configuration management, and reporting can be repurposed for new studies with only changes to configuration files and study-specific dictionaries. This level of reuse allowed implementation time for NBDC to be reduced from approximately 6 months for NBDS to just over 2 months, even as the complexity and data sources increased. The combination of configuration-driven adaptability, rigorous modularization, automated QC workflows, and open-source extensibility makes the framework a flexible and scalable solution for large, complex, and multiorganizational research environments.

In building the framework and the data system implemented using the framework, several design principles were followed with the goal of a nimble and reusable system that could be sustainably used in a myriad of different studies and implementation environments with minimal modification. The FAIR principles are detailed in Wilkinson et al (2016) [18], the principle of loose coupling, and the omission of a formal component model are the principles that allowed us to develop our framework with sufficient flexibility [19,20]. These design principles allow for a sustainable implementation, as the only changes and maintenance required on the system are dictated by updates to the open-source components used by the system.

The FAIR principles for data management suggest that contemporary data resources, tools, vocabularies, and infrastructures should exhibit the qualities required to encourage discovery and reuse [18]. Our framework follows these design principles in multiple ways. First, for the data to be findable, we keep documentation in the form of data dictionaries and enforce a persistent identifier across multiple studies and data systems for similar data types (eg, primary identifier for a blood donation is the same across studies). Second, for the data to be accessible, we built and manage an abstracted database frontend web application as part of the framework to lower the skill floor of stakeholder access to the data. Through user authentication and granular access control, the use of a web-based front-end application also contributes to data governance [13]. Third, for the data system to be interoperable, we have designed our data system to be agnostic with respect to operating systems and database engines, and to be able to connect to multiple other software platforms. Lastly, for the data to be reusable, we prioritized bidirectional compatibility and stability, wherein the data collected and managed in any 1 study should be linkable to and available for use in analyses of future data, as appropriate [20], while complying with applicable ethical and IRB requirements.

Two further guiding design principles we followed were those of loose coupling and the omission of a formal component model from the framework [19]. The omission of a formal component model was motivated by the fact that research is often not a linear endeavor, with changes to the process happening regularly as priorities shift. This is especially true for our work on the NBDS and NBDC programs. For this reason, modules were separated through workflows: data ingestion, management, and reporting. Each of these workflows can have multiple components, depending on the complexity of the study, but each workflow maintains its function, and modular code can be added or removed depending on the need.

On loose coupling, we took inspiration from ethology and decided that our framework should be agnostic with respect to operating system, relational database system of choice, and deployment method [20,36]. By choosing to develop our data system in R, and by omitting any functions that are dependent on the operating system, we allowed for the code to be run on any operating system that could run R. This also means that the system could be deployed on on-premises or cloud infrastructure. In our case, on-premises deployment was chosen, but in future studies, we plan to make use of cloud deployments,

XSL•FO
RenderX

depending on cost, data security, and cross-organization access requirements.

While adaptability and data throughput capacity are important in any large-scale research data infrastructure, for human participants' research we consider data integrity and the protection of sensitive personal information of participants to be of equal importance. Data integrity for the system is maintained through appropriate QC steps at every stage of data processing, along with the appropriate redundancies. For example, a data submission might be put through 2 QC steps, in parallel or sequentially. Automated or semiautomated handling of data submissions further supports data integrity, since rejected submissions that failed QC checks are sent back to submitters, along with a report providing detailed information on quality problems, for review, correction, and resubmission. For the protection of sensitive personal information, steps are taken to minimize the risk of a data breach by encrypting data in transit and at rest, making use of appropriately managed infrastructure (including vulnerability monitoring and an appropriate patching schedule), while also minimizing the collection and storage of personally identifying information. This ensures that even in the event of a data breach, the risk to donors or participants is minimized. An example of a practice we used to minimize the identifiability of research participants was maintaining dates of birth only at the month level and avoiding storage of multiple indirect identifiers. Further examples may include storing geographic information in formats that encode larger areas, such as 3-digit rather than 5-digit ZIP codes.

Conclusions

With life sciences research becoming increasingly reliant on large, interconnected datasets and databases, the need for adaptable, reusable, and scalable methods to manage and curate data at an organizational or multiorganizational level will also increase [14,37]. Our work shows that open-source software, along with community-led software ecosystems, can be used to meet this need and provide superior functionality and flexibility to costly proprietary data management platforms.

With the rapid advancements in technology, bioinformatics and data management have become crucial to life sciences research. This is especially true for large-scale studies addressing significant scientific and societal issues, such as the COVID-19 pandemic. The described framework was built on the need for a nimble and abstracted data management system for multicenter studies. This was achieved by using modular components, and a single project-wide configuration file in YAML format sets all module settings, including QC processes, data dictionaries, credentials, and database connection strings. These features streamline data quality control, data flow, and data formatting, while lowering the effort required to implement changes and lowering the skill floor required to manage a system that usually requires a skilled data manager.

Acknowledgments

We would like to express our deepest gratitude to our colleagues at Creative Testing Solutions, Westat, the Centers for Disease Control and Prevention (CDC), and Vitalant Research Institute for their encouragement and constructive feedback. We also extend our appreciation to the CDC for their financial support, which made this research possible. Generative artificial intelligence was not used in the preparation of this paper.

Funding

This study was funded by the CDC. The funder had no involvement in this study's design, data collection, analysis, interpretation, or writing of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data dictionary used for the first iteration of this study.
[XLSX File, 31 KB - ojphi_v18i1e81119_app1.xlsx]

Multimedia Appendix 2
Data dictionary used for the repeat donor cohort iteration of the serosurveillance study.
[XLSX File, 31 KB - ojphi_v18i1e81119_app2.xlsx]

Multimedia Appendix 3
Data submission guidelines used to train collaborators during the data submission process.
[XLSX File, 141 KB - ojphi_v18i1e81119_app3.xlsx]

References

XSL•FO
RenderX

1. Pal S, Mondal S, Das G, Khatua S, Ghosh Z. Big data in biology: the hope and present-day challenges in it. Gene Rep 2020 Dec;21:100869. [doi: 10.1016/j.genrep.2020.100869]
2. Iqbal N, Kumar P. From data science to bioscience: emerging era of bioinformatics applications, tools and challenges. Procedia Comput Sci 2023;218:1516-1528. [doi: 10.1016/j.procs.2023.01.130]
3. Liu Y, Chen Y, Han L. Bioinformatics: advancing biomedical discovery and innovation in the era of big data and artificial intelligence. TIME 2023;1(1):100012. [doi: 10.59717/j.xinn-med.2023.100012]
4. Boles NC, Stone T, Bergeron C, Kiehl TR. Big data access and infrastructure for modern biology: case studies in data repository utility. Ann N Y Acad Sci 2017 Jan;1387(1):112-123. [doi: 10.1111/nyas.13281] [Medline: 27801987]
5. Attwood TK, Blackford S, Brazas MD, Davies A, Schneider MV. A global perspective on evolving bioinformatics and data science training needs. Brief Bioinformatics 2019 Mar 25;20(2):398-404. [doi: 10.1093/bib/bbx100]
6. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019 Dec;6(1):54. [doi: 10.1186/s40537-019-0217-0]
7. Stone M, Di Germanio C, Wright DJ, et al. Use of US blood donors for national serosurveillance of severe acute respiratory syndrome coronavirus 2 antibodies: basis for an expanded national donor serosurveillance program. Clin Infect Dis 2022 Mar 9;74(5):871-881. [doi: 10.1093/cid/ciab537] [Medline: 34111244]
8. Adalja AA, Toner E, Inglesby TV. Priorities for the US health community responding to COVID-19. JAMA 2020 Apr 14;323(14):1343-1344. [doi: 10.1001/jama.2020.3413] [Medline: 32125355]
9. Rahman S, Rahman MM, Miah M, et al. COVID-19 reinfections among naturally infected and vaccinated individuals. Sci Rep 2022 Jan 26;12(1):1438. [doi: 10.1038/s41598-022-05325-5] [Medline: 35082344]
10. Cohen C, Pulliam J. COVID-19 infection, reinfection, and the transition to endemicity. Lancet 2023 Mar;401(10379):798-800. [doi: 10.1016/S0140-6736(22)02634-4]
11. Jones JM, Manrique IM, Stone MS, et al. Estimates of SARS-CoV-2 seroprevalence and incidence of primary SARS-CoV-2 infections among blood donors, by COVID-19 vaccination status - United States, April 2021-September 2022. MMWR Morb Mortal Wkly Rep 2023 Jun 2;72(22):601-605. [doi: 10.15585/mmwr.mm7222a3] [Medline: 37262007]
12. Fink RV, Fisher L, Sulaeman H, et al. How do we…form and coordinate a national serosurvey of SARS-CoV-2 within the blood collection industry? Transfusion 2022 Jul;62(7):1321-1333. [doi: 10.1111/trf.16943] [Medline: 35607854]
13. Leonelli S. Global data for local science: assessing the scale of data infrastructures in biological and biomedical research. Biosocieties 2013 Dec;8(4):449-465. [doi: 10.1057/biosoc.2013.23]
14. Harrow J, Drysdale R, Smith A, Repo S, Lanfear J, Blomberg N. ELIXIR: providing a sustainable infrastructure for life science data at European scale. Bioinformatics 2021 Aug 25;37(16):2506-2511. [doi: 10.1093/bioinformatics/btab481] [Medline: 34175941]
15. Fortunato L, Galassi M. The case for free and open source software in research and scholarship. Phil Trans R Soc A 2021 May 17;379(2197). [doi: 10.1098/rsta.2020.0079]
16. R: a language and environment for statistical computing. URL: https://www.R-project.org [accessed 2026-03-21]
17. Python. URL: https://python.org [accessed 2026-03-21]
18. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018. [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
19. Pressman RS. Software Engineering: A Practitioner's Approach, 7th edition: McGraw-Hill/Higher Education; 2010. URL: https://www.mlsu.ac.in/econtents/16_EBOOK-7th_ed_software_engineering_a_practitioners_approach_by_roger_s._pressman_.pdf [accessed 2026-03-21]
20. Bauch A, Adamczyk I, Buczek P, et al. openBIS: a flexible framework for managing and analyzing complex data in biology research. BMC Bioinformatics 2011 Dec 8;12(1):468. [doi: 10.1186/1471-2105-12-468] [Medline: 22151573]
21. hasanregius/workflow-based-information-management. GitHub. URL: https://github.com/hasanregius/workflow-based-information-management [accessed 2026-03-21]
22. Rudolph K, Schubert M. Box: write reusable, composable and modular R code. CRAN: Package box. 2025 Nov 28. URL: https://cran.r-project.org/web/packages/box/index.html [accessed 2026-03-21]
23. Ben-Kiki O, Evans C, döt Net I. YAML Ain't Markup Language (YAMLTM) Version 12. 2009 Oct 1. URL: https://yaml.org/spec/1.2.1/ [accessed 2026-03-21]
24. Curl: a modern and flexible web client for R. CRAN: Package curl. 2025 Aug 19. URL: https://cran.r-project.org/web/packages/curl/index.html [accessed 2026-03-21]
25. stenevang/sftp. GitHub. URL: https://github.com/stenevang/sftp [accessed 2026-03-21]
26. RCurl: general network (HTTP/FTP/...) client interface for R. CRAN: Package RCurl. 2025 Mar 22. URL: https://cran.r-project.org/web/packages/RCurl/index.html [accessed 2026-03-21]
27. Di Germanio C, Deng X, Balasko B, et al. P‐CB‐3 | anti‐spike and nucleocapsid antibody dynamics following SARS‐CoV‐2 infection and vaccination: implications for sourcing COVID‐19 convalescent plasma. Transfusion 2023 Oct;63(S5):148A. [doi: 10.1111/trf.186_17554]
28. Grebe E, Stone M, Spencer BR, et al. Detection of nucleocapsid antibodies associated with primary SARS-CoV-2 infection in unvaccinated and vaccinated blood donors. Emerg Infect Dis 2024 Aug;30(8):1621-1630. [doi: 10.3201/eid3008.240659] [Medline: 38981189]

XSL•FO
RenderX

29. Stone M, Grebe E, Sulaeman H, et al. Evaluation of commercially available high-throughput SARS-CoV-2 serologic assays for serosurveillance and related applications. Emerg Infect Dis 2022 Mar;28(3):672-683. [doi: 10.3201/eid2803.211885] [Medline: 35202525]
30. Sulaeman H, Grebe E, Dave H, et al. Evaluation of Ortho VITROS and Roche Elecsys S and NC immunoassays for SARS-CoV-2 serosurveillance applications. In: Chao DY, editor. Microbiol Spectr 2023 Aug 17;11(4):e0323422. [doi: 10.1128/spectrum.03234-22] [Medline: 37347180]
31. Naeem M, Jamal T, Diaz-Martinez J, et al. Trends and future perspective challenges in big data. In: Pan JS, Balas VE, Chen CM, editors. Advances in Intelligent Data Analysis and Applications: Springer, Singapore; 2022:309-325. [doi: 10.1007/978-981-16-5036-9_30]
32. Gupta A, Kumar S, Kumar A. Big data in bioinformatics and computational biology: basic insights. In: Mandal S, editor. Methods Mol Biol 2024;2719:153-166. [doi: 10.1007/978-1-0716-3461-5_9] [Medline: 37803117]
33. Xia J, Wang J, Niu S. Research challenges and opportunities for using big data in global change biology. Glob Chang Biol 2020 Nov;26(11):6040-6061. [doi: 10.1111/gcb.15317] [Medline: 32799353]
34. Awotunde JB, Oluwabukonla S, Chakraborty C, Bhoi AK, Ajamu GJ. Application of artificial intelligence and big data for fighting COVID-19 pandemic. In: Hassan SA, Mohamed AW, Alnowibet KA, editors. Decision Sciences for COVID-19: Learning Through Case Studies: Springer International Publishing; 2022:3-26. [doi: 10.1007/978-3-030-87019-5_1]
35. Nathan R, Monk CT, Arlinghaus R, et al. Big-data approaches lead to an increased understanding of the ecology of animal movement. Science 2022 Feb 18;375(6582):eabg1780. [doi: 10.1126/science.abg1780] [Medline: 35175823]
36. Glassman RB. Persistence and loose coupling in living systems. Syst Res 1973 Mar;18(2):83-98. [doi: 10.1002/bs.3830180202]
37. Kasprzyk A. BioMart: driving a paradigm change in biological data management. Database (Oxford) 2011;2011:bar049. [doi: 10.1093/database/bar049] [Medline: 22083790]

Abbreviations

CDC: Centers for Disease Control and Prevention
DCC: data coordinating center
ETL: extract, transform, load
FAIR: Findable, Accessible, Interoperable, Reusable
NBDC: Nationwide Blood Donor Cohort
NBDS: Nationwide Blood Donor Seroprevalence
QC: quality control
sFTP: secure file transfer protocol
VRI: Vitalant Research Institute
WIM: workflow-based information management
YAML: yet another markup language

Edited by E Mensah, N Zuidhof; submitted 22.Jul.2025; peer-reviewed by JK Punzalan, MS Mekbib; revised version received 23.Jan.2026; accepted 18.Feb.2026; published 20.Apr.2026.

Please cite as:
Sulaeman H, Stone M, Bruhn R, Zurita K, Nguyen A, Chiang V, Jones JM, Deng X, Custer B, Busch M, Grebe E
Workflow-Based Information Management Framework for Multicenter Research Studies: Design and Development
Online J Public Health Inform 2026;18:e81119
URL: https://ojphi.jmir.org/2026/1/e81119
doi:10.2196/81119

© Hasan Sulaeman, Mars Stone, Roberta Bruhn, Karla Zurita, Anh Nguyen, Vincent Chiang, Jefferson Michael Jones, Xutao Deng, Brian Custer, Michael Busch, Eduard Grebe. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 20.Apr.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Readability of Patient-Facing Information in Public Reporting System: Cross-Sectional Evaluation of the City of Helsinki Website

Lanmei Du, MSc; Reima Suomi, PhD

Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, Rehtorinpellonkatu 3, Turku, Finland

Corresponding Author:
Lanmei Du, MSc
Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, Rehtorinpellonkatu 3, Turku, Finland

Abstract

Background: Public reporting websites have become increasingly important for improving health care transparency and supporting informed decision-making. However, little empirical evidence exists regarding whether the readability of these websites adequately meets users' information needs, particularly in the context of digital health and Finland's ongoing social welfare and health care reform.

Objective: This study aimed to evaluate the readability of health information on the City of Helsinki public reporting website.

Methods: A cross-sectional website evaluation was conducted using the English-language health care section of the City of Helsinki public reporting website. A total of 429 web pages were collected, among which 246 unique textual descriptions were identified for readability analysis. Website text was automatically extracted using Python and evaluated using 7 established readability indexes: Flesch Reading Ease Score (FRES), Flesch-Kincaid Grade Level (FKGL), Gunning fog index (GFI), Simple Measure of Gobbledygook (SMOG), Coleman-Liau index (CLI), automated readability index (ARI), and Dale-Chall readability score (DCRS). Descriptive statistics were used to summarize textual characteristics and readability outcomes.

Results: All readability indexes consistently demonstrated that the website content exceeded the readability levels recommended for patient education materials. The mean FRES was 42.33 (SD 11.12), indicating difficult-to-read content. The mean FKGL score was 12.03 (SD 2.15), corresponding to a 12th-grade reading level, whereas the mean GFI score of 14.43 (SD 2.43) indicated undergraduate-level reading ability. Similar levels of linguistic complexity were observed across the SMOG (mean 13.69, SD 1.70), CLI (mean 13.61, SD 1.87), ARI (mean 13.67, SD 2.56), and DCRS (mean 11.49, SD 1.02). Considerable variability in readability across web pages suggested inconsistent linguistic accessibility within the website.

Conclusions: The City of Helsinki public reporting website could be improved by enhancing readability and user-centered communication. These improvements may facilitate users' understanding of health care information and support informed health care decision-making. The findings provide practical implications for the design and evaluation of future digital health care public reporting systems.

(Online J Public Health Inform 2026;18:e93241) doi:10.2196/93241

KEYWORDS

public reporting; readability; patient education materials; website evaluation; health literacy; health care transparency; digital health

Introduction

With the digitalization of health care, public reporting websites have become an important channel for increasing transparency and supporting informed decision-making [1]. Public reporting provides accessible information about the performance of hospitals, physicians, and health care providers, including treatment outcomes, infection rates, and case complexity [2-5]. Public reporting information enables patients to compare organization providers and may support their choices as part of shared decision-making within patient-physician comanagement [6-9].

However, the usefulness of public reporting depends not only on whether information is publicly available but also on whether users can understand and interpret it correctly [10-13]. Medical information differs from many other types of information because misunderstanding may have direct and potentially irreversible consequences for patients [14,15]. Therefore, health care information should be presented with particular attention to accuracy, clarity, comprehensibility, and unbiased reporting [16-22]. Clear explanations, appropriate contextualization, and understandable presentation of quality indicators can help users with different levels of health literacy interpret public reporting information and avoid misjudgment [23-27].

XSL•FO
RenderX

This issue can also be understood as a problem of representation. As illustrated by Plato's allegory of the cave, what people perceive may be a representation rather than reality itself [28]. Similarly, developers and users of health care information systems interact with coded and presented representations of health care reality rather than with health care reality directly [29,30]. What patients encounter on a public reporting website is therefore not health care quality itself but a particular representation of that quality through indicators, numbers, labels, explanations, and textual descriptions [18]. From a human-computer interaction perspective, Norman [31] described a "gulf" between users' conceptualization of a task and the interface presented to them. How information is presented and communicated through an interface may therefore influence users' ability to understand and use it [32,33].

Information quality is therefore an important consideration in the design and evaluation of public reporting websites [34-36]. Information is generally understood as processed data that are meaningful and valuable for current or future actions and decisions [37]. Higher information quality has been associated with greater user satisfaction and more effective system use [38]. In particular, the representational quality of information concerns how information is presented to users, including its simplicity, presentation style, and comprehensibility [39]. For public reporting websites, these characteristics are especially important because quality indicators and other performance measures may be difficult for patients to interpret when they are highly technical or insufficiently explained [40-44]. In addition, narrative or other unstructured language may require greater processing effort, providing context that can supplement and support the interpretation of data-driven information [45,46].

One important aspect of information presentation is readability. Readability affects the effort required to process written information and may influence whether users can successfully obtain and understand the information they need [47]. Although previous research has examined public reporting in relation to patient choice, patient satisfaction, transparency, and health care quality [48-52], empirical research assessing the actual medical information presented on public reporting websites remains limited [53-56]. In particular, little is known about whether patient-facing public reporting information is sufficiently readable to support users' information needs in the context of digital health and shared decision-making.

This issue is particularly relevant in Finland, where the social welfare and health care reform emphasizes freedom of choice and the development of public reporting systems to support informed choices [57]. As public reporting systems and digital health services become increasingly important sources of health care information, evaluating the readability of patient-facing public reporting content can help determine whether these systems effectively communicate information to the public [58-60]. Therefore, this study asked the following question: How readable is patient-facing public reporting information? To address this question, we evaluated the readability of the City of Helsinki public reporting website.

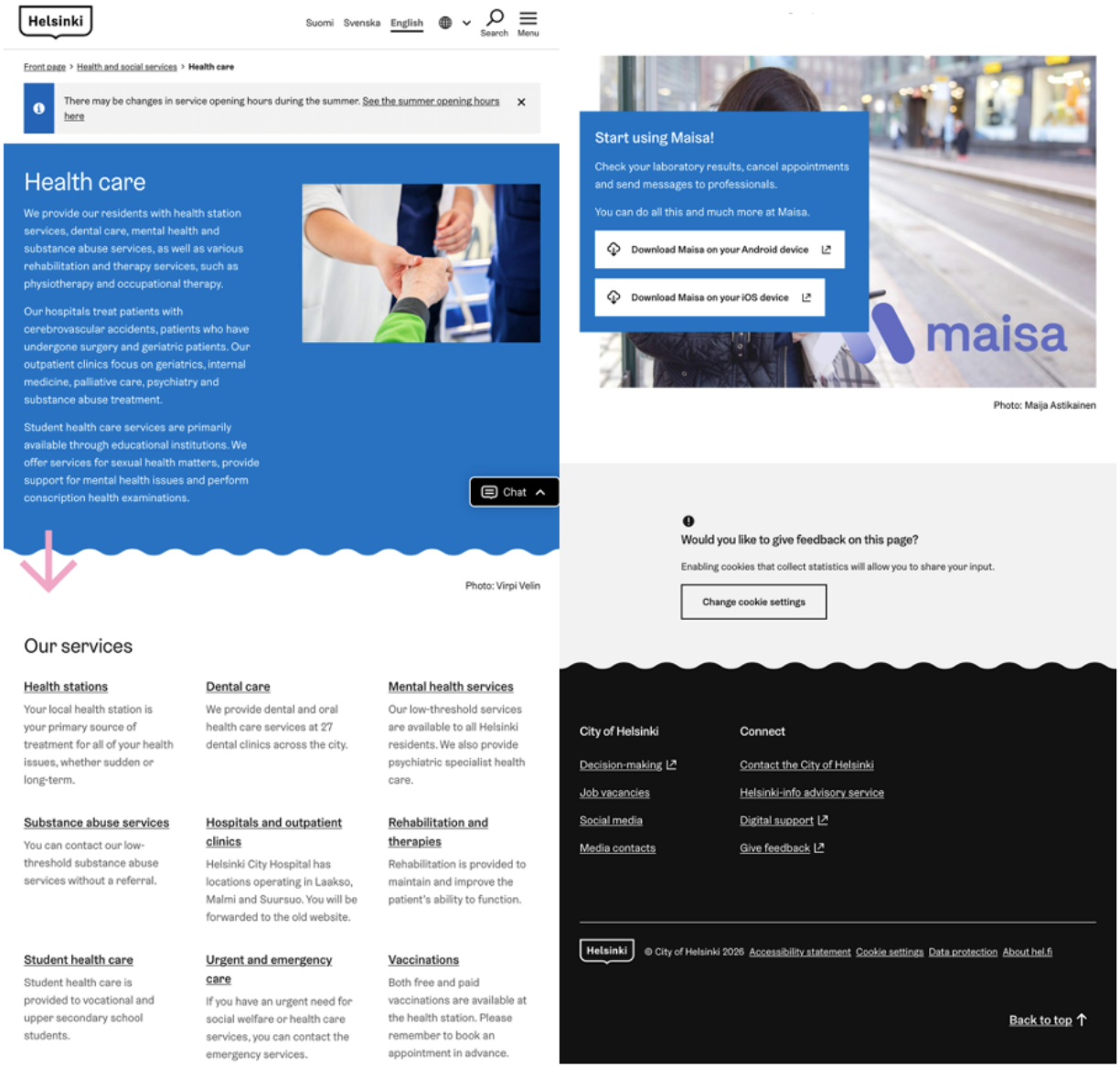
Methods

Study Design

This study used a cross-sectional website evaluation design. A total of 429 web pages were collected from the health care section of the City of Helsinki website at the page level. Accordingly, the analysis was restricted to the health care section of the City of Helsinki website, which was selected as the primary object of analysis. Representative screenshots of the City of Helsinki health care website are provided in Figure 1 to illustrate the website interface and page layout.

XSL•FO
RenderX

Figure 1. Home page of the health care section of the City of Helsinki website as an example of a health care website, presented in 2 consecutive sections because of the web page length.



In this study, public reporting websites were conceptualized as textual information resources. It should be noted that a substantial proportion of the City of Helsinki web pages consisted of short headings and hyperlink labels.

Internal web pages within the City of Helsinki health care website were included in the analysis. Hyperlink labels appearing on these web pages were retained as part of the visible text because they constitute information presented to users. In contrast, hyperlinks directing users to external websites were recorded but not followed or included as analytical units.

Furthermore, because readers do not naturally filter out redundant or navigational content when accessing information on web pages, all textual elements were retained for analysis. Therefore, no minimum text length threshold (eg, excluding texts containing fewer than 100 words) was applied, allowing all visible English-language textual content to be included in the readability analysis.

All eligible visible textual content within each web page was aggregated into a single web page–level text corpus before readability assessment. Therefore, headings and hyperlink labels were not evaluated as independent short text units. Readability was then assessed at the web page level using 7 established readability indexes.

This study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies. The completed STROBE checklist is provided in Checklist 1.

Data Collection

All websites were assessed in November 2025. Data collection was conducted using Python (version 3.9.6; Python Software Foundation), with packages including *Requests*, *Beautiful Soup*, *Fake-useragent*, *urllib.parse*, and *openpyxl*.

First, all textual content was scraped from the website. To ensure accuracy and minimize errors, author LD reviewed the dataset.

XSL•FO
RenderX

The study then established a Microsoft Excel sheet containing the URLs of each website (Multimedia Appendix 1).

Subsequently, all 429 extracted URLs, together with the textual content from each web page, were exported to a Microsoft Excel file. The web page was defined as the primary unit of sampling. Subsequent verification of the URLs revealed that some represented anchor URLs pointing to different sections within the same web page. Duplicate textual descriptions across different numbers were identified, revealing that 183 URLs contained duplicated text (Multimedia Appendix 2).

Nevertheless, website users encounter these web pages individually during navigation and do not manually remove duplicate content. After identifying duplicated textual content and identical readability scores, 246 unique textual records were retained for statistical analysis.

The text associated with each URL was recorded as the corresponding description. Only English-language web pages were included because the readability indexes used in this study were validated primarily for English-language text and rely on English-specific linguistic features such as word length, syllable count, and familiar word lists. In addition, PDF documents, images, and other nontextual content (eg, external links) were excluded from the readability analysis.

Readability Assessment

Following data extraction, web page records with identical textual content and identical readability scores were identified and deduplicated using a custom Python script (Multimedia Appendix 3). Readability indexes and text characteristics were then summarized using descriptive statistics.

Readability-related text characteristics were extracted for each web page, including the total number of characters, words, and sentences; mean sentence length (words per sentence); total syllables; number of polysyllabic words; and the proportion of polysyllabic words among all words (Multimedia Appendix 4).

Readability is commonly defined as the ease with which a text can be read and understood [61]. In health literacy and patient education research, it is often described more precisely as “the study of matching reader and text,” so that the reading level of the material corresponds to the reader’s ability to read with understanding [62].

Readability was assessed as an indicator of linguistic complexity that may influence users’ cognitive processing of public reporting information. Existing conceptual frameworks for public reporting websites include readability alongside dimensions such as accuracy, completeness, design, usability, and accessibility [63]. Automated website assessment frameworks operationalize these dimensions across domains such as accessibility, content, marketing, and technology [64].

Numerical scores were generated automatically using predefined computational algorithms. No subjective human scoring was involved. Before the readability assessment, web page text was preprocessed by removing noncontent elements, including scripts, HTML markup, stylesheets, content from external websites, and PDF documents where applicable.

Within the Nibbler website assessment framework, readability is evaluated as an indicator within the “Accessibility” and “Content” categories [65-67]. Given that the conceptualization and operationalization of readability vary across disciplines, this study used 7 established readability measures to capture complementary aspects of text difficulty. These measures differ in their underlying linguistic features and weighting schemes, allowing the consistency of readability assessments to be examined across alternative operationalizations.

The primary outcomes were the 7 readability indexes (FRES, FKGL, GFI, SMOG, CLI, ARI, and DCRS). Descriptive statistics, including the mean, SD, and minimum and maximum values, were then calculated and are presented in the Results section.

Using the readability measures as shown in Table 1, readability was assessed using the Python Textstat library, which was used to calculate 7 established readability measures: the Flesch Reading Ease Score (FRES) [68], Flesch-Kincaid Grade Level (FKGL) [65], Gunning fog index (GFI) [69,70], Simple Measure of Gobbledygook (SMOG) [41], Coleman-Liau index (CLI) [71], automated readability index (ARI) [72], and Dale-Chall readability score (DCRS) [73]. These indexes estimate the reading difficulty of health-related materials and have been extensively applied in previous studies evaluating online health information [65].

XSL•FO
RenderX

Table . Standard formulas and descriptions of the readability measures used in this study.

Readability measures	Focus	Scoring basis	Formula
FRES[a]	Scores range from 0 to 100, with higher scores indicating easier readability.	Sentence length; syllables per word	206.835 − (84.6 × ASW[b]) − (1.015 × ASL[c])
FKGL[d]	Estimates the approximate US grade level required to understand a text, with higher scores indicating greater reading difficulty.	Sentence length; syllables per word	(0.39 × ASL) + (11.8 × ASW) − 15.59
GFI[e]	Estimates the years of formal education required to understand a text, emphasizing sentence length and complex words.	Sentence length; complex words	0.4 × (ASL + [(C[f]/W[g]) × 100])
SMOG[h]	Estimates the approximate grade level required to understand a text based on the number of PWs[i].	PWs; sentence count	1.0430 × √PW + 3.1291
CLI[j]	Estimates the approximate grade level required to comprehend a text using characters and sentence-based measures.	Character-based word length and sentence-based features	(0.0588 × L[k]) − (0.296 × S[l]) − 15.8
ARI[m]	Estimates the approximate US grade level required to understand a text using character length and sentence length rather than syllable counts.	Sentence length; characters per word	4.71 × (characters/words) + (0.5 × ASL) − 21.43
DCRS[n]	Estimates reading difficulty based on sentence length and the proportion of words not included in a predefined list of familiar words.	Sentence length; DWs[o]	0.1579 × (DW/W × 100) + (0.0496 × ASL) + 3.6365[p]

[a]FRES: Flesch Reading Ease Score.

[b]ASW: average syllables per word.

[c]ASL: average sentence length (number of words per sentence).

[d]FKGL: Flesch-Kincaid Grade Level.

[e]GFI: Gunning fog index.

[f]C: number of complex words after excluding proper nouns, words made polysyllabic by adding “ed” or “es,” and compound words composed of simpler words.

[g]W: total number of words.

[h]SMOG: Simple Measure of Gobbledygook.

[i]PW: polysyllabic word (words with ≥3 syllables).

[j]CLI: Coleman-Liau index.

[k]L: average number of letters per 100 words.

[l]S: average number of sentences per 100 words.

[m]ARI: automated readability index.

[n]DCRS: Dale-Chall readability score.

[o]DW: difficult word not included in the Dale-Chall familiar word list.

[p]The constant 3.6365 is added when the proportion of difficult words exceeds 5%.

Although some measures rely on overlapping variables, they differ in how these variables are weighted and interpreted. The FRES estimates reading ease, whereas the FKGL, GFI, SMOG, and CLI provide approximate US grade-level estimates. The ARI similarly provides a grade-level estimate based on character and sentence length, whereas the DCRS incorporates lexical familiarity through a predefined familiar word list. Thus, the 7 measures were treated as complementary indicators of readability rather than as 7 independent outcomes.

For interpretation, higher FRES scores indicate greater reading ease, whereas higher scores on the other grade-level or difficulty-oriented measures indicate greater reading difficulty. The US Department of Health and Human Services sixth-grade reading level recommendation corresponds to a FRES score of 80 or higher and FKGL, GFI, SMOG, and CLI scores of 6.9 or lower [74-76].

Ethical Considerations

This study analyzed only publicly available website content and did not involve human participants, personal data, or any intervention. Therefore, formal ethics approval was not required. According to the Finnish National Board on Research Integrity

XSL•FO
RenderX

guidelines, ethical review is generally not required for research based solely on publicly available data [77].

Results

A total of 246 unique textual records from the health care section of the City of Helsinki website were included in the analysis. Table 2 shows the descriptive statistics of the readability measures for patient-facing website content, whereas the descriptive statistics of the textual characteristics of the website content are provided in Multimedia Appendix 4.

Table . Descriptive statistics of the readability measures for patient-facing website content.[a]

Readability measures	Mean (SD)	Range
Flesch Reading Ease Score	42.33 (11.12)	5.53-63.13
Flesch-Kincaid Grade Level	12.03 (2.15)	8.41-17.99
Gunning fog index	14.43 (2.43)	10.25-21.75
Simple Measure of Gobbledygook	13.69 (1.70)	10.54-19.05
Coleman-Liau index	13.61 (1.87)	9.60-19.84
Automated readability index	13.67 (2.56)	9.21-22.07
Dale-Chall readability score	11.49 (1.02)	8.85-14.43

[a]Higher values for the Flesch Reading Ease Score indicate easier-to-read content, whereas higher values for the Flesch-Kincaid Grade Level, Gunning fog index, Simple Measure of Gobbledygook, Coleman-Liau index, automated readability index, and Dale-Chall readability score indicate more difficult-to-read content.

The mean FRES score was 42.33 (SD 11.12), corresponding to a “difficult” readability level according to the established Flesch scale [68]. The mean FKGL score was 12.03 (SD 2.15), corresponding to approximately a 12th-grade reading level and substantially exceeding the recommended range of 6 to 8 [78]. Similarly, the mean GFI, SMOG, CLI, and ARI scores were 14.43 (SD 2.43), 13.69 (SD 1.70), 13.61 (SD 1.87), and 13.67 (SD 2.56), respectively, indicating reading levels substantially above the recommended range. The mean DCRS was 11.49 (SD 1.02), also indicating considerable reading difficulty. Overall, all 7 readability measures consistently indicated that the patient-facing public reporting web pages were written at relatively high reading difficulty levels.

Discussion

Principal Findings

This study found that patient-facing public reporting web pages consistently exceeded the readability levels recommended for patient education materials, suggesting that transparency alone does not guarantee that health care information is understandable or usable by patients. Although public reporting websites make health care performance information publicly available, the linguistic complexity of the content indicates a potential challenge for patient-facing communication. This distinction is particularly important when patients seek public reporting information to compare health care providers, understand health care performance, or inform their health care decisions. When information is publicly available but difficult to understand, formal transparency may not translate into meaningful access or effective use.

The magnitude of the observed readability difficulty further underscores this concern. The mean GFI score of 14.43 corresponds to a readability level estimated at approximately 14 years of formal education, or a college-level reading requirement, whereas the SD of 2.43 indicates moderate variation in readability across web pages [79]. The FRES results similarly indicated a college-level reading requirement. These findings substantially exceed the grade 6 to 8 readability level recommended by the American Medical Association and the National Institutes of Health for written health materials, particularly those intended for individuals with limited health literacy [80,81]. Thus, when patients turn to public reporting websites to obtain information relevant to their health care choices, the relatively high reading difficulty observed in this study may represent a potential challenge for comprehension and subsequent use of the information [82]. Rather than simply increasing the amount of information available to patients, patient-centered public reporting should therefore consider whether users can understand and meaningfully apply the information provided.

The consistency of the findings across the 7 readability measures further strengthens this interpretation. Although these measures differ in their underlying variables, weighting schemes, and interpretations, they consistently indicated a relatively high level of reading difficulty. FRES and FKGL both incorporate sentence length and syllables per word but differ in how their scores are interpreted, whereas the GFI emphasizes complex words, the SMOG focuses on polysyllabic words, the CLI and ARI rely on character-based features, and the DCRS incorporates lexical familiarity. The convergence of these measures suggests that the observed difficulty is not attributable to a single readability formula but reflects a broader pattern of linguistic complexity in patient-facing public reporting content. The relatively high GFI scores indicate substantial linguistic complexity across many web pages, potentially including specialized medical terminology that may be difficult for lay users, such as patients and family members, to process [82]. Despite differences in their underlying formulas and weighting

XSL•FO
RenderX

schemes, the 7 indexes therefore provide converging evidence supporting the robustness of the overall findings.

These findings also have implications beyond the immediate readability of patient-facing materials. Health literacy involves individuals' ability to access, understand, evaluate, and use health information, whereas data literacy encompasses the ability to interpret and meaningfully use data in a given context [83]. In increasingly digital health environments, effective use of public reporting information may require both forms of literacy. Within public reporting systems, this process involves multiple groups, including health care professionals who generate or interpret performance information, developers who translate underlying data into digital interfaces, and patients and other users who ultimately interpret and apply the information presented to them. Public reporting websites therefore function not merely as channels for patient education but also as shared information environments in which health literacy and data literacy are enacted. Ensuring that complex performance information is communicated in accessible language may consequently support more equitable participation in the interpretation and use of health care data.

High-quality information is valuable only when it can be readily understood and meaningfully used by individuals with different levels of health literacy [84]. There is previous evidence suggesting that individuals with higher educational attainment tend to prefer access to raw data, whereas those with lower educational attainment rely more heavily on intuitive representations, such as color coding and graphical displays [85-87]. The present study did not examine differences in users' educational backgrounds or health literacy. Instead, the high readability difficulty observed in the public reporting web pages draws attention to the linguistic accessibility of patient-facing information and the need to consider variation in users' information needs and capabilities when designing and communicating public reporting information. Future research could examine whether combining appropriately simplified textual information with visualizations and other user-oriented presentation formats can improve users' comprehension, interpretation, and use of public reporting information.

Previous studies have emphasized that narrative-based communication may strengthen users' understanding and interpretation of public reporting systems by improving the transparency, credibility, and perceived relevance of publicly reported health care information [88-90]. Building on this literature, our findings suggest that contextualized communication alone may not be sufficient if the language used to convey the information remains difficult for the intended audience to understand. Readability should therefore be considered complementary to narrative and other communication strategies aimed at improving transparency and credibility. In the context of patient-provider shared decision-making, public reporting systems may be more effective when information is not only transparent and contextually explained but also presented in a language that patients can readily understand and apply to their health care decisions.

Policy Implications for Patient-Centered Public Reporting Information

Increasing transparency of health care performance reporting should be regarded not only as making health care performance data publicly available but also as ensuring that these data are presented in ways that facilitate public understanding. Accordingly, government agencies responsible for public reporting websites may consider incorporating readability evaluation into routine website quality assessment. Automated readability assessment tools could be integrated into website maintenance workflows to identify pages requiring revision and improve the consistency of patient-facing information.

These considerations may become applicable as digital health information is incorporated into AI-enabled health information services. Publicly available web pages may serve as information sources for systems that retrieve, summarize, or generate health-related information. Improving the quality, consistency, and readability of public reporting web pages may therefore have implications beyond direct patient use by improving the quality of information available to downstream digital and AI-enabled health information services. However, this potential benefit should be interpreted cautiously as further research is needed to determine whether improvements in web page readability translate into greater accuracy, interpretability, or reliability of AI-generated health information.

Practical Implications

Improving the quality of public reporting websites should be considered alongside the broader processes through which health care information is stored, structured, validated, and published. The public reporting data presented on websites represent only a subset of the information maintained within health care databases [91,92]. Website data generally consist of structured and unstructured content, with semistructured formats such as XML functioning as an intermediary layer for data validation, exchange, and standardized representation [93]. Optimizing XML-based data structures during database development can improve the accuracy and consistency of public reporting data, thereby enhancing transparency [94].

At the same time, redevelopment of existing information systems typically requires substantial time, financial investment, and organizational resources, making ecosystem redesign impractical in many health care settings [95]. Under these constraints, optimizing intermediate data structures (eg, XML) represents a practical strategy for improving key data processing workflows and enhancing the consistency and standardization of publicly reported information, thereby supporting more transparent and reliable public reporting. However, improvements in the underlying data infrastructure should complement rather than replace efforts to improve the readability and user-centered presentation of information. Ultimately, effective public reporting requires both reliable underlying data and communication practices that enable patients, health care professionals, and other users to understand and meaningfully use those data.

Previous research suggests that the way in which information is structured and visually presented can influence how users

XSL•FO
RenderX

attend to and respond to publicly reported health care information [96-100]. Websites commonly use visual elements such as icons, symbols, and color to communicate information, whereas the quantity and presentation of information may also affect users' ability to process and use it [101-103]. Studies of public reporting and health care information presentation have further examined users' responses to quality ratings, visual displays, and other forms of information presentation [96-98]. These findings highlight the importance of considering not only what information is presented but also how it is structured and communicated to users. We provide an initial examination of the readability of the underlying information presented on patient-facing public reporting web pages, highlighting a potential issue at the data and content level that warrants further investigation.

From a practical perspective, data development processes could therefore incorporate annotations and contextual information within the underlying code to better accommodate users. Such annotations could provide greater clarity regarding the meaning and interpretation of data elements, which may contribute to more readable and understandable information when the data are subsequently presented on web pages. Similarly, web page design should consider how colors, symbols, and other visual elements are likely to be interpreted by users. Selecting and applying these elements in ways that are intuitive and readily interpretable may further support the accessibility and usability of patient-facing public reporting information.

Limitations

This study has several limitations. First, the readability analysis was performed only on web pages with textual content. Although no minimum text length threshold was applied, manual inspection indicated that web pages containing fewer than 100 words were mainly composed of headings, navigation labels, short notices, or introductions. Therefore, these short text segments were retained for descriptive statistical analysis. Future website research should perform word segmentation and apply appropriate text exclusion criteria once sufficient textual data are available to ensure more robust readability assessment.

Second, the readability results were based on automated algorithmic scores rather than direct user testing. During the manual review of all extracted texts, we observed that the public reporting system presented relatively limited performance data but included a dedicated feedback form for reporting website-related issues. As user heterogeneity in health literacy, educational level, language proficiency, and familiarity with health care terminology exists [42,104], future studies could investigate dashboard-based public reporting systems with user testing that integrate audiovisual elements to enhance information presentation [105,106].

Finally, the data were obtained from the English-language health-related pages of a single municipal website in Finland. Therefore, the findings primarily reflect this specific website and its linguistic and operational context, and caution is warranted when transferring the findings to other public reporting environments. Future research could examine public reporting websites across different organizational, operational, and language settings. In particular, comparative analyses of Finnish- and English-language content could determine whether readability differs across language versions of the same website.

Conclusions

This study evaluated the readability of the City of Helsinki public reporting website from the perspective of health care information design. The findings consistently demonstrated that the website content exceeded the readability levels recommended for patient education materials and exhibited considerable variability across web pages.

These findings suggest that the effectiveness of public reporting depends not only on the transparency of health care data but also on how such information is encoded, organized, and communicated to end users.

Improving readability and narrative-based communication may enhance users' understanding of public reporting information and, ultimately, promote more informed health care decision-making. Future research should further examine multilingual public reporting websites and investigate how different information presentation strategies influence users' comprehension and engagement across diverse health care contexts.

Acknowledgments

The authors are grateful to the anonymous reviewers for their comments on the key concept clarification between websites, AI, and individuals. The authors thank the editors for their insights on the structure and clarification between AI, automation tools, and user heterogeneity of the manuscript. During the preparation of this manuscript, generative AI (ChatGPT; OpenAI) was used solely for translation and language polishing purposes. All outputs generated by the AI were carefully reviewed and verified by the authors to ensure accuracy and appropriateness. The AI was not involved in study design; data collection, analysis, and interpretation; or writing of the scientific content of the manuscript.

Funding

LD was financially supported by the China Scholarship Council (grant 202306770017). The funder had no role in the study design; data collection, analysis, and interpretation; or manuscript writing.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study and the code used for data collection and analysis are available in the GitHub repository [107].

Authors' Contributions

Conceptualization: LD, RS
Data curation: LD
Formal analysis: LD
Investigation: LD, RS
Methodology: LD
Resources: RS
Software: LD
Supervision: RS
Validation: LD
Visualization: LD
Writing—original draft: LD
Writing—review and editing: RS

Conflicts of Interest

None declared.

Multimedia Appendix 1
URLs with description.
[XLSX File, 225 KB - ojphi_v18i1e93241_app1.xlsx]

Multimedia Appendix 2
Readability results before deduplication.
[XLSX File, 153 KB - ojphi_v18i1e93241_app2.xlsx]

Multimedia Appendix 3
Readability results after deduplication.
[XLSX File, 64 KB - ojphi_v18i1e93241_app3.xlsx]

Multimedia Appendix 4
Descriptive statistics of the linguistic characteristics of the website content.
[DOCX File, 17 KB - ojphi_v18i1e93241_app4.docx]

Checklist 1
STROBE checklist.
[DOCX File, 33 KB - ojphi_v18i1e93241_app5.docx]

References

1. Cordova-Pozo K. Digitalization in healthcare and health data reporting: opportunities to reduce error and inequality of healthcare delivery. Int J Community Based Nurs Midwifery 2025 Apr;13(2):161-163. [doi: 10.30476/ijcbnm.2025.106424.2761] [Medline: 40322056]
2. Cacace M, Geraedts M, Berger E. Public reporting as a quality strategy. In: Busse R, Klazinga N, Panteli D, Quentin W, editors. Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies: European Observatory on Health Systems and Policies; 2019:331-355.
3. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. JAMA 2000 Apr 12;283(14):1866-1874. [doi: 10.1001/jama.283.14.1866] [Medline: 10770149]
4. Behrendt K, Groene O. Mechanisms and effects of public reporting of surgeon outcomes: a systematic review of the literature. Health Policy 2016 Oct;120(10):1151-1161. [doi: 10.1016/j.healthpol.2016.08.003] [Medline: 27638232]
5. Prang KH, Maritz R, Sabanovic H, Dunt D, Kelaher M. Mechanisms and impact of public reporting on physicians and hospitals' performance: a systematic review (2000–2020). PLoS One 2021;16(2):e0247297. [doi: 10.1371/journal.pone.0247297] [Medline: 33626055]

XSL•FO
RenderX

6. Galletta M, Piazza MF, Meloni SL, et al. Patient involvement in shared decision-making: do patients rate physicians and nurses differently? Int J Environ Res Public Health 2022 Oct 31;19(21):14229. [doi: 10.3390/ijerph192114229] [Medline: 36361109]
7. Shay LA, Lafata JE. Understanding patient perceptions of shared decision making. Patient Educ Couns 2014 Sep;96(3):295-301. [doi: 10.1016/j.pec.2014.07.017] [Medline: 25097150]
8. Coulter A, Entwistle V, Gilbert D. Sharing decisions with patients: is the information good enough? BMJ 1999 Jan 30;318(7179):318-322. [doi: 10.1136/bmj.318.7179.318] [Medline: 9924064]
9. Turner AM, Osterhage KP, Taylor JO, Hartzler AL, Demiris G. A closer look at health information seeking by older adults and involved family and friends: design considerations for health information technologies. AMIA Annu Symp Proc 2018;2018:1036-1045. [Medline: 30815147]
10. Abdelaziz S, Garfield S, Neves AL, et al. What are the unintended patient safety consequences of healthcare technologies? A qualitative study among patients, carers and healthcare providers. BMJ Open 2024 Nov 27;14(11):e089026. [doi: 10.1136/bmjopen-2024-089026] [Medline: 39608995]
11. Smith MA, Gigot M, Harburn A, et al. Insights into measuring health disparities using electronic health records from a statewide network of health systems: a case study. J Clin Transl Sci 2023;7(1):e54. [doi: 10.1017/cts.2022.521] [Medline: 37008604]
12. Seong D, Jung S, Bae S, Chung J, Son DS, Yi BK. Fast Healthcare Interoperability Resources (FHIR)-based quality information exchange for clinical next-generation sequencing genomic testing: implementation study. J Med Internet Res 2021 Apr 28;23(4):e26261. [doi: 10.2196/26261] [Medline: 33908889]
13. Mishra NK, Duke J, Lenert L, Karki S. Public health reporting and outbreak response: synergies with evolving clinical standards for interoperability. J Am Med Inform Assoc 2020 Jul 1;27(7):1136-1138. [doi: 10.1093/jamia/ocaa059] [Medline: 32692844]
14. Flores G, Laws MB, Mayo SJ, et al. Errors in medical interpretation and their potential clinical consequences in pediatric encounters. Pediatrics 2003 Jan;111(1):6-14. [doi: 10.1542/peds.111.1.6] [Medline: 12509547]
15. Riva G. The sociocognitive psychology of computer-mediated communication: the present and future of technology-based interactions. Cyberpsychol Behav 2002 Dec;5(6):581-598. [doi: 10.1089/109493102321018222] [Medline: 12556121]
16. Johnson C, Peterson SK, Turley JP, et al. Strategies to improve healthcare websites. AMIA Annu Symp Proc 2006;2006:969. [Medline: 17238588]
17. Abdel-Wahab N, Rai D, Siddhanamatha H, Dodeja A, Suarez-Almazor ME, Lopez-Olivo MA. A comprehensive scoping review to identify standards for the development of health information resources on the internet. PLoS One 2019;14(6):e0218342. [doi: 10.1371/journal.pone.0218342] [Medline: 31220126]
18. Sapin M, Ehlig D, Geissler A, Vogel J. Public reporting in five health care areas: a comparative content analysis across nine countries. Health Policy 2025 Feb;152:105222. [doi: 10.1016/j.healthpol.2024.105222] [Medline: 39700634]
19. Saghafian S, Hopp WJ. Can public reporting cure healthcare? The role of quality transparency in improving patient–provider alignment. Oper Res 2020;68(1):71-92. [doi: 10.1287/opre.2019.1868]
20. Frank LB, Murphy ST, Chatterjee JS, Moran MB, Baezconde-Garbanati L. Telling stories, saving lives: creating narrative health messages. Health Commun 2015;30(2):154-163. [doi: 10.1080/10410236.2014.974126] [Medline: 25470440]
21. Chalil Madathil K, Greenstein JS. Designing comprehensible healthcare public reports: an investigation of the use of narratives and tests of quality metrics to support healthcare public report sensemaking. Appl Ergon 2021 Sep;95:103452. [doi: 10.1016/j.apergo.2021.103452] [Medline: 33957305]
22. Glässel A, Tyebally Fang M, Hendriks MJ, Biller-Andorno N, The Swiss DIPEx Research Group. Using narratives to improve health literacy – an ethical and public health perspective. In: Barrett DH, Ortmann LW, Larson SA, editors. Narrative Ethics in Public Health: The Value of Stories: Springer; 2022:169-180. [doi: 10.1007/978-3-030-92080-7_12]
23. Loy M, Kowalsky R. Narrative medicine: the power of shared stories to enhance inclusive clinical care, clinician well-being, and medical education. Perm J 2024 Jun 14;28(2):93-101. [doi: 10.7812/TPP/23.116] [Medline: 38225914]
24. Mazor KM, Dodd KS, Kunches L. Communicating hospital infection data to the public: a study of consumer responses and preferences. Am J Med Qual 2009;24(2):108-115. [doi: 10.1177/1062860608330827] [Medline: 19237552]
25. Lagu T, Greaves F. From public to social reporting of hospital quality. J Gen Intern Med 2015 Oct;30(10):1397-1399. [doi: 10.1007/s11606-015-3327-1] [Medline: 25933624]
26. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. JAMA 2005 Mar 9;293(10):1239-1244. [doi: 10.1001/jama.293.10.1239] [Medline: 15755946]
27. Sander U, Emmert M, Dickel J, Meszmer N, Kolb B. Information presentation features and comprehensibility of hospital report cards: design analysis and online survey among users. J Med Internet Res 2015 Mar 16;17(3):e68. [doi: 10.2196/jmir.3414] [Medline: 25782186]
28. Bloom A. The Republic of Plato, 2nd edition: Basic Books; 1991.
29. Zaharias G. What is narrative-based medicine? Narrative-based medicine 1. Can Fam Physician 2018 Mar;64(3):176-180. [Medline: 29540381]
30. Liu D, Wang X, Pan F, Xu Y, Yang P, Rao K. Web-based infectious disease reporting using XML forms. Int J Med Inform 2008 Sep;77(9):630-640. [doi: 10.1016/j.ijmedinf.2007.10.011] [Medline: 18060833]

XSL•FO
RenderX

31. Norman D. The Design of Everyday Things: Basic Books; 2013.
32. Damman OC, Hendriks M, Rademakers J, Spreeuwenberg P, Delnoij DM, Groenewegen PP. Consumers' interpretation and use of comparative information on the quality of health care: the effect of presentation approaches. Health Expect 2012 Jun;15(2):197-211. [doi: 10.1111/j.1369-7625.2011.00671.x] [Medline: 21615637]
33. 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]
34. DeLone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2003 Apr;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
35. Petter S, DeLone W, McLean E. Measuring information systems success: models, dimensions, measures, and interrelationships. Eur J Inf Syst 2008 Jun;17:236-263. [doi: 10.1057/ejis.2008.15]
36. March ST, Smith GF. Design and natural science research on information technology. Decis Support Syst 1995 Dec;15(4):251-266. [doi: 10.1016/0167-9236(94)00041-2]
37. Davis GB, Olson MH. Management Information Systems: Conceptual Foundations, Structure, and Development: McGraw-Hill; 1996.
38. Kalankesh LR, Nasiry Z, Fein RA, Damanabi S. Factors influencing user satisfaction with information systems: a systematic review. Galen Med J 2020;9:e1686. [doi: 10.31661/gmj.v9i0.1686] [Medline: 34466567]
39. Laumer S, Maier C, Weitzel T. Information quality, user satisfaction, and the manifestation of workarounds: a qualitative and quantitative study of enterprise content management system users. Eur J Inf Syst 2017;26(4):333-360. [doi: 10.1057/s41303-016-0029-7]
40. Christianson JB, Volmar KM, Alexander J, Scanlon DP. A report card on provider report cards: current status of the health care transparency movement. J Gen Intern Med 2010 Nov;25(11):1235-1241. [doi: 10.1007/s11606-010-1438-2] [Medline: 20625849]
41. Duncanson E, Davies CE, Muthuramalingam S, et al. Patient perspectives of center-specific reporting in kidney failure care: an Australian qualitative study. Kidney Int Rep 2024;9(4):843-852. [doi: 10.1016/j.ekir.2024.01.001] [Medline: 38765598]
42. Shaller D, Kanouse DE, Schlesinger M. Context-based strategies for engaging consumers with public reports about health care providers. Med Care Res Rev 2014 Oct;71(5 Suppl):17S-37S. [doi: 10.1177/1077558713493118] [Medline: 23819945]
43. Helminski D, Sussman JB, Pfeiffer PN, et al. Development, implementation, and evaluation methods for dashboards in health care: scoping review. JMIR Med Inform 2024 Dec 10;12:e59828. [doi: 10.2196/59828] [Medline: 39656991]
44. Helminski D, Kurlander JE, Renji AD, et al. Dashboards in health care settings: protocol for a scoping review. JMIR Res Protoc 2022 Mar 2;11(3):e34894. [doi: 10.2196/34894] [Medline: 35234650]
45. Peng H. Quantifying narrative data in qualitative analysis: an abductive approach and interactive process [Abstract]. Presented at: 2022 Annual Meeting of Academy of Management; Aug 5-9, 2022. [doi: 10.5465/AMBPP.2022.14953abstract]
46. Lieblich A, Tuval-Mashiach R, Zilber T. Narrative Research: Reading, Analysis, and Interpretation: SAGE Publications; 1998.
47. Ahmed FZ, Noblin A, Courtelyou K, Tynes G. Evaluating the readability, accessibility, and quality of online health information on opioid use disorder: a comparative analysis of government and non-government websites. Adv Health Inf Sci Pract 2026;2(1):eSVXO9846. [doi: 10.63116/SVXO9846] [Medline: 41743651]
48. Totten AM, Miake-Lye IM, Vaiana ME, et al. Public Presentation of Health System or Facility Data about Quality and Safety: A Systematic Review: Department of Veterans Affairs (US); 2011. [Medline: 22206114]
49. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Med Care 2003 Jan;41(1 Suppl):I30-I38. [doi: 10.1097/00005650-200301001-00004] [Medline: 12544814]
50. Shintani Smith S, Cheng BT, Kern RC, Cameron KA, Micco AG. Publicly reported patient satisfaction scores in academic otolaryngology departments. Laryngoscope 2021 Oct;131(10):2204-2210. [doi: 10.1002/lary.29557] [Medline: 33830507]
51. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946. [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
52. Timofeyev Y, Dremova O, Jakovljevic M. The impact of transparency constraints on the efficiency of the Russian healthcare system: systematic literature review. J Med Econ 2023;26(1):95-109. [doi: 10.1080/13696998.2022.2160608] [Medline: 36537319]
53. Ye J, Hai J, Song J, Wang Z. The role of artificial intelligence in the application of the integrated electronic health records and patient-generated health data. medRxiv. Preprint posted online on May 3, 2024. [doi: 10.1101/2024.05.01.24306690]
54. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J 2021 Jul;8(2):e188-e194. [doi: 10.7861/fhj.2021-0095] [Medline: 34286183]
55. Fernandes Prabhu D, Gurupur V, Stone A, Trader E. Integrating artificial intelligence, electronic health records, and wearables for predictive, patient-centered decision support in healthcare. Healthcare (Basel) 2025 Oct 30;13(21):2753. [doi: 10.3390/healthcare13212753] [Medline: 41228120]
56. Rehburg F, Graefe A, Hübner M, Thun S. How interoperability can enable artificial intelligence in clinical applications. Stud Health Technol Inform 2024 Aug 22;316:596-600. [doi: 10.3233/SHTI240485] [Medline: 39176813]

XSL•FO
RenderX

57. Collective agreement for social and health care personnel or SOTE agreement. SuPer. URL: https://www.superliitto.fi/en/working-life/employment/collective-agreements/sote-agreement/ [accessed 2026-07-24]
58. Tan DJ, Ko TK, Fan KS. The readability and quality of web-based patient information on nasopharyngeal carcinoma: quantitative content analysis. JMIR Form Res 2023 Nov 27;7:e47762. [doi: 10.2196/47762] [Medline: 38010802]
59. Ozduran E, Büyükçoban S. Evaluating the readability, quality and reliability of online patient education materials on post-COVID pain. PeerJ 2022;10:e13686. [doi: 10.7717/peerj.13686] [Medline: 35880220]
60. McGrath L, Millar BC, Moore JE. Using plain language to communicate with clinical trials participants: comparison of readability calculators. Contemp Clin Trials 2022 Dec;123:106995. [doi: 10.1016/j.cct.2022.106995] [Medline: 36347454]
61. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. Clin Orthop Relat Res 2010 Oct;468(10):2572-2580. [doi: 10.1007/s11999-010-1380-y] [Medline: 20496023]
62. Okuhara T, Furukawa E, Okada H, Kiuchi T. Readability of online and offline written health information: a protocol of a systematic review of systematic reviews. BMJ Open 2024 Dec 23;14(12):e079756. [doi: 10.1136/bmjopen-2023-079756] [Medline: 39719278]
63. Siddhanamatha HR, Heung E, Lopez-Olivo ML, et al. Quality assessment of websites providing educational content for patients with rheumatoid arthritis. Semin Arthritis Rheum 2017 Jun;46(6):715-723. [doi: 10.1016/j.semarthrit.2017.01.006] [Medline: 28258768]
64. Gale JJ, Black KC, Calvano JD, et al. An analysis of US academic medical center websites: usability study. J Med Internet Res 2021 Dec 21;23(12):e27750. [doi: 10.2196/27750] [Medline: 34932015]
65. Kincaid JP, Fishburne RP Jr, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. : Institute for Simulation and Training; 1975.
66. Silktide. URL: https://silktide.com/platform/ [accessed 2026-09-29]
67. Huerta TR, Hefner JL, Ford EW, McAlearney AS, Menachemi N. Hospital website rankings in the United States: expanding benchmarks and standards for effective consumer engagement. J Med Internet Res 2014 Feb 25;16(2):e64. [doi: 10.2196/jmir.3054] [Medline: 24568892]
68. Flesch R. A readability formula in practice. Elem Engl 1948;25(6):344-351 [FREE Full text]
69. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013;9(5):503-516. [doi: 10.1016/j.sapharm.2012.05.009] [Medline: 22835706]
70. Mac O, Ayre J, Bell K, McCaffery K, Muscat DM. Comparison of readability scores for written health information across formulas using automated vs manual measures. JAMA Netw Open 2022 Dec 1;5(12):e2246051. [doi: 10.1001/jamanetworkopen.2022.46051] [Medline: 36508219]
71. Coleman M, Liau TL. A computer readability formula designed for machine scoring. J Appl Psychol 1975;60(2):283-284. [doi: 10.1037/h0076540]
72. Mc Carthy A, Taylor C. SUFE and the internet: are healthcare information websites accessible to parents? BMJ Paediatr Open 2020 Oct;4(1):e000782. [doi: 10.1136/bmjpo-2020-000782] [Medline: 33094174]
73. Koo K, Yap RL. How readable is BPH treatment information on the internet? Assessing barriers to literacy in prostate health. Am J Mens Health 2017 Mar;11(2):300-307. [doi: 10.1177/1557988316680935] [Medline: 27903952]
74. Wrigley Kelly NE, Murray KE, McCarthy C, O'Shea DB. An objective analysis of quality and readability of online information on COVID-19. Health Technol (Berl) 2021;11(5):1093-1099. [doi: 10.1007/s12553-021-00574-2] [Medline: 34189011]
75. Kim P, Eng TR, Deering MJ, Maxfield A. Published criteria for evaluating health related web sites: review. BMJ 1999 Mar 6;318(7184):647-649. [doi: 10.1136/bmj.318.7184.647] [Medline: 10066209]
76. Fundingsland ELJ, Fike J, Calvano J, Beach J, Lai D, He S. Methodological guidelines for systematic assessments of health care websites using web analytics: tutorial. J Med Internet Res 2022;24(4):e28291. [doi: 10.2196/28291] [Medline: 35436216]
77. The ethical principles of research with human participants and ethical review in the human sciences in Finland. Finnish National Board on Research Integrity TENK Publications. 2019.
78. Nattam A, Vithala T, Wu TC, et al. Assessing the readability of online patient education materials in obstetrics and gynecology using traditional measures: comparative analysis and limitations. J Med Internet Res 2023 Aug 30;25:e46346. [doi: 10.2196/46346] [Medline: 37647115]
79. Varli B, Cetindag EN, Koyuncu Demir K, Coban U, Islamova G, Dokmeci F. Evaluation of readability levels of online patient education materials for female pelvic floor disorders. Medicine (Baltimore) 2023 Dec 29;102(52):e36636. [doi: 10.1097/MD.0000000000036636] [Medline: 38206737]
80. Eltorai AE, Ghanian S, Adams CA Jr, Born CT, Daniels AH. Readability of patient education materials on the American Association for Surgery of Trauma website. Arch Trauma Res 2014 Jun;3(2):e18161. [doi: 10.5812/atr.18161] [Medline: 25147778]
81. Para A, Thelmo F, Rynecki ND, et al. Evaluating the readability of online patient education materials related to orthopedic oncology. Orthopedics 2021 Jan 1;44(1):38-42. [doi: 10.3928/01477447-20201012-04] [Medline: 33141234]

XSL•FO
RenderX

82. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
83. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80. [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
84. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int 2000 Sep 1;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
85. Liu C, Tang Y, Wang D, Zhang X. The effect of public reporting presentation on patients' decision making: an experimental survey in Yunan Province, China. Medicine (Baltimore) 2017;96(24):e7203. [doi: 10.1097/MD.0000000000007203] [Medline: 28614266]
86. Emmert M, Schlesinger M. Hospital quality reporting in the United States: does report card design and incorporation of patient narrative comments affect hospital choice? Health Serv Res 2017 Jun;52(3):933-958. [doi: 10.1111/1475-6773.12519] [Medline: 27324087]
87. Shi Y, Scanlon DP, Bhandari N, Christianson JB. Is anyone paying attention to physician report cards? The impact of increased availability on consumers' awareness and use of physician quality information. Health Serv Res 2017 Aug;52(4):1570-1589. [doi: 10.1111/1475-6773.12540] [Medline: 27468943]
88. Charon R. What to do with stories: the sciences of narrative medicine. Can Fam Physician 2007 Aug;53(8):1265-1267. [Medline: 17872831]
89. Whitehead A, Woods A, Atkinson S, Macnaughton J, Richards J, editors. The Edinburgh Companion to the Critical Medical Humanities: Edinburgh University Press; 2016.
90. Li J. Public reporting and consumer demand in the home health sector. Med Care Res Rev 2024 Apr;81(2):122-132. [doi: 10.1177/10775587231221852] [Medline: 38235538]
91. Annis A, Reaves C, Sender J, Bumpus S. Health-related data sources accessible to health researchers from the US government: mapping review. J Med Internet Res 2023 Apr 27;25:e43802. [doi: 10.2196/43802] [Medline: 37103987]
92. Weintraub WS, Garratt KN. Public reporting II: state of the art-current public reporting in cardiovascular medicine. Circulation 2017 May 9;135(19):1772-1774. [doi: 10.1161/CIRCULATIONAHA.117.026553] [Medline: 28483827]
93. Structured vs unstructured content: why it matters. Agility CMS. 2025. URL: https://agilitycms.com/blog/structured-vs-unstructured-content-why-it-matters [accessed 2026-09-29]
94. Codd EF. A relational model of data for large shared data banks. Commun ACM 1970 Jun;13(6):377-387. [doi: 10.1145/362384.362685]
95. Kahn CE Jr, de la Cruz NB. Extensible markup language (XML) in health care: integration of structured reporting and decision support. Proc AMIA Symp 1998:725-729. [Medline: 9929314]
96. Leiter RE, Pu CT, Mazzola E, et al. Engaging hospices in quality measurement and improvement: early experiences of a large integrated health care system. J Pain Symptom Manage 2020 Oct;60(4):866-873.e4. [doi: 10.1016/j.jpainsymman.2020.05.027] [Medline: 32512046]
97. Perraillon MC, Konetzka RT, He D, Werner RM. Consumer response to composite ratings of nursing home quality. Am J Health Econ 2019;5(2):165-190. [doi: 10.1162/ajhe_a_00115] [Medline: 31579236]
98. Temkin-Greener H, Mao Y, McGarry B. Online customer reviews of assisted living communities: association with community, county, and state factors. J Am Med Dir Assoc 2023 Jun;24(6):841-845.e3. [doi: 10.1016/j.jamda.2023.02.007] [Medline: 36934775]
99. Yılmaz NG, Timmermans DR, Van Weert JC, Damman OC. Breast cancer patients' visual attention to information in hospital report cards: an eye-tracking study on differences between younger and older female patients. Health Informatics J 2023;29(1):14604582231155279. [doi: 10.1177/14604582231155279] [Medline: 36840473]
100. Theis RP, Stanford JC, Goodman JR, Duke LL, Shenkman EA. Defining "quality" from the patient's perspective: findings from focus groups with Medicaid beneficiaries and implications for public reporting. Health Expect 2017 Jun;20(3):395-406. [doi: 10.1111/hex.12466] [Medline: 27124419]
101. Lee T, Lee-Geiller S, Lee BK. Are pictures worth a thousand words? The effect of information presentation type on citizen perceptions of government websites. Gov Inf Q 2020 Jul;37(3):101482. [doi: 10.1016/j.giq.2020.101482]
102. Slovic P. Toward understanding and improving decisions. In: Science, Technology, and the Modern Navy: Thirtieth Anniversary, 1946-1976: Office of Naval Research; 1976.
103. Keller KL, Staelin R. Effects of quality and quantity of information on decision effectiveness. J Consum Res 1987 Sep;14(2):200-213. [doi: 10.1086/209106]
104. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187. [doi: 10.2196/jmir.2702] [Medline: 23985220]
105. Bos VL, Jansen T, Klazinga NS, Kringos DS. Development and actionability of the Dutch COVID-19 dashboard: descriptive assessment and expert appraisal study. JMIR Public Health Surveill 2021 Oct 12;7(10):e31161. [doi: 10.2196/31161] [Medline: 34543229]

XSL•FO
RenderX

106. Barbazza E, Ivanković D, Wang S, et al. 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 Aug 6;23(8):e30200. [doi: 10.2196/30200] [Medline: 34280120]
107. Dulanmei/web-data-readability-pr-system. GitHub. URL: https://github.com/dulanmei/web-data-readability-pr-system [accessed 2026-09-30]

Abbreviations

ARI: automated readability index
CLI: Coleman-Liau index
DCRS: Dale-Chall readability score
FKGL: Flesch-Kincaid Grade Level
FRES: Flesch Reading Ease Score
GFI: Gunning fog index
SMOG: Simple Measure of Gobbledygook
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L MacNeill; submitted 10.Feb.2026; peer-reviewed by J Mino-Ayala, Z Su; revised version received 19.Sep.2026; accepted 21.Sep.2026; published 08.Oct.2026.

Please cite as:
Du L, Suomi R
The Readability of Patient-Facing Information in Public Reporting System: Cross-Sectional Evaluation of the City of Helsinki Website
Online J Public Health Inform 2026;18:e93241
URL: https://ojphi.jmir.org/2026/1/e93241
doi:10.2196/93241

© Lanmei Du, Reima Suomi. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 8.Oct.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Immunization Data Systems to Track Zero-Dose Children in Bangladesh: Mixed Methods Exploratory Study

Samiha Yunus[1], MPH; Prem Singh[2], MDCHA; Alan Noble John[2], MPH; Elizabeth Oliveras[3], ScD; Md Jasim Uddin[4], PhD; Nurul Alam[4], PhD; Md Wazed Ali[4], MSS; Hemel Das[4], MSc; Chris Morgan[3], PhD

[1]Jhpiego, Dhaka, Bangladesh
[2]Jhpiego, New Delhi, India
[3]Jhpiego, Maryland, MD, United States
[4]Health Systems and Population Studies Division, icddr,b, Dhaka, Bangladesh

Corresponding Author:
Samiha Yunus, MPH
Jhpiego
Plot:10, Level:10, Road 90 Gulshan 2
Dhaka, 1212
Bangladesh
Phone: 880 01730434951
Email: samiha.yunus@jhpiego.org

Abstract

Background: Zero-dose (ZD) children—those who receive no routine vaccines—remain highly vulnerable, with global evidence showing that they are concentrated in marginalized and hard-to-reach settings. In Bangladesh, stagnant full-immunization coverage, population mobility, and inconsistencies in administrative data complicate accurate estimation of ZD and underimmunized children. Gaps in the completeness, timeliness, and accuracy of routine Expanded Programme on Immunization (EPI) information systems further limit the ability to identify and target these populations.

Objective: This study aimed to assess the completeness, timeliness, and accuracy of Bangladesh's immunization data systems and to identify challenges and strategies for integrating ZD analysis into routine EPI information systems.

Methods: We used a mixed methods exploratory descriptive qualitative approach comprising a desk review and stakeholder interviews conducted from February 2024 to April 2024. The desk review assessed program documents, gray literature, and published studies to map routine immunization data sources and their relevance for ZD analysis. Thirteen purposively selected stakeholders, including EPI and District Health Information Software 2 (DHIS2) managers, were interviewed using semistructured interview guides to gather insights on data availability, reporting practices, and system challenges. Information from both steps was synthesized using a structured checklist to evaluate completeness, timeliness, accuracy, and feasibility of ZD tracking across data sources. Data were analyzed manually using a deductive thematic approach.

Results: Ten routine immunization data sources were identified, with DHIS2, the Rapid Convenience Monitoring (RCM) tool, and the EPI Coverage Evaluation Survey (CES) emerging as the primary systems for ZD tracking. Although most sources captured child-level data, the use of unique identifiers and household-level variables was inconsistent. Data quality and timeliness varied: DHIS2 faced denominator inaccuracies, CES and Bangladesh Demographic and Health Survey (BDHS) provided high-quality survey estimates, and RCM delivered real-time data from hard-to-reach areas. Multiple systems supported ZD identification, although each had limitations related to data completeness, sampling, geographic granularity, or operational constraints.

Conclusions: Immunization data systems in Bangladesh are fragmented, with limited ZD tracking and predominantly district-level reporting. DHIS2, RCM, and Health and Demographic Surveillance System (HDSS) are the most promising platforms for routine ZD monitoring, offering nationwide integration, real-time field data, and detailed local analysis. Harmonizing these systems and combining data from multiple sources can improve targeted interventions, subnational monitoring, and equitable vaccine coverage.

(Online J Public Health Inform 2026;18:e88951) doi:10.2196/88951

KEYWORDS

immunization; zero-dose; vaccine; data system; children; Bangladesh

XSL•FO
RenderX

Introduction

Although immunization is one of the most effective and cost-efficient public health interventions, approximately 15 million children worldwide remained unvaccinated in 2023 [1]. These children, classified as zero-dose (ZD) children, received none of the routine childhood vaccines, leaving them highly vulnerable to preventable and potentially life-threatening diseases [1]. Systemic inequalities in immunization systems contribute to the lack of access to necessary vaccinations [2]. In low- and middle-income countries (LMICs), children who miss their first dose of the diphtheria, tetanus, and pertussis (DTP1) vaccination account for nearly half of all vaccine-preventable deaths [3]. Evidence reported that ZD children are usually concentrated in urban poor, remote rural, or conflict-affected areas, with more than 45% of the 9.7 million children residing in such geographic settings [4-6]. Evidence from a Country Learning Hub (CLH) rapid assessment report corroborates global trends to some extent in the context of Bangladesh [7]. The report highlights the heightened vulnerability of migrant and internally displaced populations, alongside specific demand-side barriers. It also points to common structural determinants observed globally, including low socioeconomic status, forced migration, homelessness, and religious or cultural marginalization [7].

In Bangladesh, coverage measured in terms of fully vaccinated children has fluctuated between 80% and 84% over the past decade, resulting in a persistently undervaccinated population of 16% to 20% [8]. Bangladesh's Expanded Programme on Immunization (EPI) must reach underprivileged groups, as unvaccinated people are still susceptible to diseases such as measles and diphtheria [9]. Achieving high vaccination coverage and precisely determining the ZD burden are made more difficult by the nation's diverse, hard-to-reach communities and urban migration [7,9]. Official estimates of DTP1, collated by the World Health Organization (WHO) and United Nations Children's Fund (UNICEF), have fluctuated between 93% and 100% over the past decade, with administrative estimates often exceeding 100% [10]. Overestimated numerators or denominators can lead to errors in ZD or underimmunized estimates. Therefore, a more accurate estimation method is provided by the rapid assessment report, which also emphasizes the necessity of resolving issues with health information systems to prioritize programs and monitor progress in ZD reduction [7].

Routine EPI information systems are foundational to immunization programs, supporting the planning, implementation, and evaluation of vaccination services. These systems routinely collect data from community and facility levels and play a critical role in identifying areas with low coverage, including ZD children, to guide targeted and equitable interventions. In Bangladesh, however, gaps in data completeness, timeliness, and accuracy may hinder the ability to effectively identify and reach ZD children. This study thus aimed to conduct a comprehensive review of Bangladesh's immunization data landscape to assess the completeness, timeliness, and accuracy of existing data sources; evaluate their strengths and limitations; and identify feasible strategies for integrating ZD analysis into routine EPI information systems.

Methods

Study Design

We adopted a mixed methods exploratory descriptive qualitative approach to assess the routine immunization data landscape, with a focus on ZD children in Bangladesh. Data collection was conducted using multiple methods, including a desk review and interviews conducted between February 2024 and April 2024. First, a rapid desk review was undertaken to identify existing immunization-related data sources and relevant stakeholders. Furthermore, we conducted interviews with stakeholders as part of the qualitative component to gather in-depth insights.

Desk Review

We conducted an in-depth desk review of program documents, gray literature, and published empirical studies related to immunization data sources in Bangladesh. The primary sources used to identify these data sources were official websites, reports, and publications. To identify the appropriate data sources, we followed multiple processes. Using keywords such as immunization, immunization data, ZD children, EPI, vaccination, coverage, hard-to-reach areas, and Bangladesh, we searched for relevant documents and websites using Google. We identified relevant websites, dashboards, literature, and reports and thoroughly reviewed each document and the respective organizations' websites. The review focused on gathering information about the characteristics of each data source, availability of data on ZD children, reporting frequency, and the strengths and limitations of each source.

Qualitative Interviews

We conducted interviews with key stakeholders involved in immunization systems and in strengthening these systems in Bangladesh. The interviews complemented the desk review. Participants were selected purposively based on their relevance to the identified immunization systems in Bangladesh. Initially, we contacted potential participants and shared the study objectives. Subsequently, interviews were conducted at times and locations convenient for the participants. For example, we interviewed EPI managers responsible for the District Health Information Software 2 (DHIS2; HISP Centre), as they are tasked with routinely updating the immunization data, including data on ZD children. The aim of these interviews was to gain valuable insights into the data sources used by vaccination programs, with a particular focus on information related to ZD children. Moreover, we obtained the necessary credentials to access the identified websites, as they were not publicly available and required proper authorization for access. A total of 13 interviews were conducted using a semistructured interview guide.

Review of Data Sources

We combined the insights gained from the interviews and the desk review to assess the selected websites and documents. Using the developed checklist, each data source was analyzed using specific themes to evaluate its capacity for ZD analysis

XSL•FO
RenderX

at the national, administrative division, and district levels. We assessed the feasibility of ZD tracking, the availability of data on ZD children, and reporting frequency and compared the sources based on these features. Furthermore, we explored the strengths and limitations associated with each data source.

Data Analysis

We conducted a descriptive and deductive thematic analysis. After the completion of the interviews, audio recordings and interview notes were promptly translated into English. Thematic analysis was conducted using the 6-phase approach developed by Braun and Clarke [11]. Using this approach, responses were categorized under key themes, including the availability of data on ZD children, data quality, timeliness, strengths, and limitations. Coding was conducted independently by 2 study team members, with regular meetings held to compare, discuss, and reconcile differences in coding. Discrepancies were resolved through discussion and consensus, with oversight from other team members. To enhance rigor, peer debriefing was conducted throughout the analysis to support interpretation and ensure consistency in theme development. Triangulation was achieved by comparing findings across the desk review and interviews. The data were analyzed manually and presented in tabular format to facilitate comparison and interpretation.

Ethical Considerations

This study was part of the CLH activities, and ethics approval for the study was obtained from the institutional review board of the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) before data collection (PR#22170). Informed written consent was obtained from all participants prior to the interviews, and the confidentiality and anonymity of the information they provided were ensured.

Results

Sources of Routine Immunization Data and Relevant Stakeholders

We identified a total of 10 immunization data sources in Bangladesh [12-23]. Among them, 5 (50%) data sources were online-based systems used primarily for routine data collection and monitoring. A total of 4 (40%) data sources were document-based, published periodically as survey or report outputs. The remaining 1 (10%) data source involved the use of the Rapid Convenience Monitoring (RCM) tool, a digital tool for household monitoring and immunization session assessment. The 3 major sources of routine immunization data in Bangladesh are DHIS2, the RCM tool, and the EPI Coverage Evaluation Survey (CES). DHIS2 is a nationally scaled, open-source digital health management information system covering all public health facilities. RCM, used through the WHO–Surveillance and Immunization Medical Officer (SIMO) network, provides real-time data from hard-to-reach and underserved areas via digital tools. CES is a periodic survey using multistage cluster sampling to evaluate immunization coverage and service quality through household interviews and record reviews. The list of identified data sources for routine immunization in Bangladesh is provided in Table 1.

Table 1. Sources of routine immunization data and relevant stakeholders.

Data sources	Data source's availability	Stakeholders consulted
District Health Information Software 2	Website	EPI[a] managers
EPI Coverage Evaluation Survey 2019	Report	WHO[b]
Routine immunization monitoring and supervision data	Report	WHO
Bangladesh Demographic and Health Survey	Demographic and Health Surveys website	EPI managers and partners
Health and Demographic Surveillance System data	Report—MATLAB	International Centre for Diarrhoeal Disease Research, Bangladesh
Effective Vaccine Management	Report	WHO and United Nations Children's Fund
Vaccine-preventable disease surveillance data	Report	Government officials
Adverse events following immunization surveillance data	Website	Directorate General of Drug Administration, Bangladesh
Digital tools for household monitoring and immunization session assessment	Tools supplied by the Ministry of Health and Family Welfare	EPI managers
Geospatial information systems overviews	Website	WHO; Bangladesh Geographic Information System Portal

[a]EPI: Expanded Programme on Immunization.

[b]WHO: World Health Organization.

Key Features of Data Sources

Most systems, including DHIS2, CES, RCM, Bangladesh Demographic and Health Survey (BDHS), and Health and Demographic Surveillance System (HDSS), collect detailed child-level information, including child identifiers, age, sex, and health card or vaccination registry details. However, unique identifiers are inconsistently used across systems; only DHIS2, BDHS, and HDSS incorporate them. Birth order is recorded solely by BDHS. BDHS is the most comprehensive source for

XSL•FO
RenderX

capturing household-level variables such as guardian's education and wealth quintile. CES and RCM also partially collect these variables, while DHIS2 and HDSS have limited coverage. CES includes questions about vaccination side effects, the number of required visits, and the source of vaccination. BDHS covers general awareness about immunization, but other systems have limited or no data on these indicators. CES uniquely captures information about hard-to-reach areas, while BDHS includes geographic coordinates of health facilities, which can support spatial mapping and planning. Only the vaccine-preventable disease (VPD) surveillance system records comprehensive operational elements such as training, logistic support, storage and cold chain capacity, adverse event following immunization (AEFI), population coverage, and referral systems. These elements are not addressed by the other systems. More specific information is available in Table 2.

XSL•FO
RenderX

Table 2. Key features of different data sources.

Data elements	District Health Information Software 2	Coverage Evaluation Survey 2019	Rapid Convenience Monitoring	Bangladesh Demographic and Health Survey 2017-2018	Health and Demographic Surveillance System	Vaccine-preventable disease surveillance
Child-level data						
Unique identifier	✓		✓	✓		
Child identifier	✓	✓	✓	✓	✓	✓
Card, registry, or history	✓	✓	✓	✓		✓
Age	✓	✓	✓	✓	✓	
Sex	✓	✓	✓	✓	✓	
Birth order				✓		
Anthropometric data	✓		✓	✓		
Guardian's education		✓		✓	✓	
Wealth quintile		✓		✓	✓	
Knowledge about vaccination						
Side effect of vaccination		✓				
Knowledge of number of visits		✓				
Source of vaccination		✓				
Vaccination awareness[a]					✓	
Health facility information						
Hard-to-reach area[b]		✓				
Geocoordinates of health facilities					✓	
Immunization services						
Training						✓
Logistic support						✓
Storage capacity						✓
Cold storage						✓
Population coverage						✓
Adverse event following immunization						✓
Referral system						✓

[a]Vaccination awareness includes which vaccine is being received, knowledge about pentavalent vaccination, how the participant learned about pentavalent vaccination, and knowledge about the next vaccination.

[b]Hard-to-reach areas are geographically difficult-to-access areas where the delivery of routine immunization services is constrained by factors such as remote locations, poor transportation, riverine (*char*) settings, and hill tract areas with limited access to health facilities.

Data Quality, Integrity, Timeliness, and ZD Tracking

The DHIS2 system faces restrictions due to denominator inaccuracies, as it relies on projected census data that may not accurately reflect actual facility catchment populations. In contrast, the EPI-CES and the BDHS are considered high-quality sources, benefiting from rigorous sampling methods and experienced survey teams. Routine immunization monitoring provides useful insights, but the effectiveness of the data depends heavily on adequate sample sizes for sessions and households, especially when targeting ZD populations. The HDSS contributes high-quality data through regularly shared summaries and accessible datasets, supporting robust analysis for both researchers and policymakers.

XSL•FO
RenderX

DHIS2 reports EPI data on a monthly or quarterly basis, depending on the health system level being monitored. The EPI-CES has a reporting frequency of 1 to 4 years, which allows in-depth analysis of national immunization coverage and access. RCM provides real-time data through tools such as Power BI (Microsoft Corp) and KOBO (Kobo Organization), enabling immediate access for EPI managers to track vaccination sessions and household monitoring. The BDHS releases data every 3 to 5 years, providing a valuable tool for long-term monitoring of immunization progress.

Several data sources contribute to identifying and understanding ZD children in Bangladesh. DHIS2 tracks key vaccinations such as Penta1, serving as a proxy for ZD identification, although limitations in data quality may affect accuracy. The EPI-CES provides detailed, disaggregated data on immunization coverage, enabling robust ZD estimates. The RCM tool identifies ZD and undervaccinated children while capturing community-level barriers to vaccine uptake. The BDHS offers nationally representative data to estimate ZD prevalence among children aged 12 to 23 months. Finally, the HDSS provides granular, area-specific data on immunization status and timeliness. Together, these sources offer a comprehensive understanding of ZD children, supporting targeted strategies to improve vaccine coverage. More specific information is available in Table 3.

Table 3. Data quality, integrity, timeliness, and zero-dose (ZD) tracking.

Data sources	Data quality	Timeliness	ZD tracking capacity	Data entry access level
District Health Information Software 2	Moderate; completeness concerns, especially for dropout tracking	Real-time data entry but often delayed reporting from facilities	Partial; potential with unique IDs but limited implementation	Facility-level data entry by health workers
Expanded Programme on Immunization Coverage Evaluation Survey	High; based on rigorous methodology and a large sample	Retrospective; typically conducted every 5 years	Yes; ZD children can be identifiable through survey questions	Trained survey teams collect and enter data through centralized survey management
Rapid Convenience Monitoring	Moderate; relies on field monitoring, with some inconsistency	Monthly to quarterly monitoring cycles	No direct tracking; purposive sampling of ZD clusters is needed	Field supervisors or the Surveillance and Immunization Medical Officer network enter data using digital tools
Bangladesh Demographic and Health Survey	High; standardized survey protocol with strong sampling	Every 3 to 5 years	Yes; allows identification based on receipt of the first dose of the diphtheria, tetanus, and pertussis vaccine	Dedicated national survey teams collect and enter data
Health and Demographic Surveillance System	High; maintained by the International Centre for Diarrhoeal Disease Research, Bangladesh, using strict protocols	Continuous, although not nationally representative	Yes; detailed child histories enable ZD estimation	Site-based, trained field staff enter longitudinal household data
Vaccine-preventable disease surveillance	Moderate; captures operational data but is not designed for detailed child tracking	Periodic; not real time	No	Facility-based reporting by clinicians and surveillance officers

Implementation-Related Strengths and Challenges

The integration of DHIS2 across Bangladesh's health system enabled monthly and quarterly tracking of immunization coverage, particularly with respect to ZD data, down to local levels, including Upazila and City Corporation facilities, thus supporting informed decision-making. The CES 2019 provided detailed data on immunization coverage through rigorous sampling, addressing denominators and incorporating factors such as vaccinated and unvaccinated children. In terms of real-time supervision, RCM facilitated the prompt detection of ZD children, focusing on potential clusters and capturing reasons for nonvaccination. Additionally, the BDHS 2017-2018 offered comprehensive, population-representative data that were useful for hypothesis generation, while HDSS captured unique factors, such as migration. Moreover, VPD surveillance systems and regulatory frameworks strengthened vaccine delivery and safety monitoring.

The constraints of various systems regarding immunization data tracking and ZD identification were notable (Table 3). DHIS2 faced challenges such as potential reporting bias, restricted multilevel data, reliance on aggregate data with limited individual-level insights, and the absence of household-level ZD tracking. The CES 2019 had several limitations, including its inability to provide subdistrict estimates, reliance on small clusters, high costs, recall bias, and a lack of sociobehavioral insights. Similarly, RCM depended on external resources, was at risk of biased estimates due to convenience sampling, and was restricted to district-level data. The BDHS 2017-2018 relied on immunization cards and lacked data on nonvaccination. HDSS was limited by aggregated data and restricted public access. Finally, VPD systems faced workforce shortages, geographic inequities, and cold chain gaps. More specific information is available in Table 4.

XSL•FO
RenderX

Table 4. Implementation-related challenges and scalability of health and immunization monitoring systems.

Data sources	Health workforce	Infrastructure	Data capture quality	Timeliness	Monitoring and supervision	Scalability
District Health Information Software 2	High workload on health workers; limited training	Poor infrastructure in some areas	Risk of reporting bias; limited individual-level data	Routine reporting (monthly or quarterly); delays in facility reporting	Limited data-use feedback loops	High; nationally implemented but needs quality strengthening
Coverage Evaluation Survey 2019	Requires trained survey teams	Strong logistical requirements	High-quality but subject to recall bias	Periodic (every 1 to 5 years); not real time	Strong centralized supervision during the survey	Low; resource intensive and costly
Rapid Convenience Monitoring	Dependence on field staff capacity	Requires digital tools and connectivity	Inconsistent reporting; potential sampling bias	Monthly to quarterly cycles	Strong reliance on supervisory effectiveness	Moderate; depends on resources and supervision
Bangladesh Demographic and Health Survey	Requires experienced national survey teams	High survey logistics requirements	High quality but subject to recall bias	Every 3 to 5 years; long intervals	Strong national survey oversight	Low; not designed for routine monitoring
Health and Demographic Surveillance System	Requires skilled field surveillance staff	Site-based infrastructure only	High-quality longitudinal data; limited generalizability	Continuous data collection	Strong site-level monitoring systems	Low to moderate; expansion is resource intensive
Vaccine-preventable disease surveillance	Clinical and surveillance staff burden	Health facility–based systems; cold chain dependency	Incomplete community capture; operational data focus	Periodic; not real time	Variable supervision across facilities	Moderate; complements other systems

Discussion

Principal Findings and Comparison With Prior Work

This study found that Bangladesh has multiple immunization data systems with complementary strengths, including routine digital reporting systems, periodic surveys, and surveillance platforms. However, no single system provides complete, timely, and consistently high-quality data for identifying and tracking ZD children. Although systems such as DHIS2, CES, BDHS, HDSS, and RCM each contribute valuable information, gaps remain in data completeness, standardization, and individual-level tracking. Timeliness also varies widely across systems, ranging from real-time reporting to survey cycles spanning multiple years. Overall, the findings highlight the need for better integration and harmonization of existing data sources to support effective ZD monitoring within routine EPI systems.

Among the identified sources, 1 major data source, the EPI-CES, primarily focuses on routine immunization services delivered through fixed-site facilities. In contrast, another significant source, the EPI Outreach Monitoring System, is concerned with immunization coverage achieved through outreach activities, particularly among hard-to-reach and underserved populations. The lack of an integrated and comprehensive national data system that simultaneously captures both facility-based and outreach-based immunization services poses a challenge for monitoring equitable coverage and planning targeted interventions [24]. This fragmentation in immunization data systems has also been noted in previous assessments, which emphasized the need for harmonized and timely data collection tools to improve vaccine coverage monitoring and program planning in LMICs, including Bangladesh [25,26].

Furthermore, we observed that multiple health and vaccination monitoring systems in Bangladesh use different approaches to data collection and immunization coverage tracking. Each system contributes uniquely to identifying ZD children and monitoring vaccination service delivery. This diversity highlights both the strengths and challenges of having multiple, often uncoordinated, data sources for immunization tracking. The disintegration of these systems may impede comprehensive coverage analysis and strategic planning; however, when effectively harmonized, they also offer valuable opportunities for designing and implementing more targeted and context-specific interventions [27,28].

The study also revealed that DHIS2 has played a crucial role in monitoring vaccination coverage at different levels of the health system, such as in Upazila Health Complexes. DHIS2 is well suited for capturing and managing health data, including ZD data. As part of Bangladesh's health information infrastructure, DHIS2 supports routine tracking of immunization data, enabling timely, data-driven decision-making. The integration of features into DHIS2 to capture ZD data has accelerated its potential. Like Bangladesh, several countries, including Mozambique, Nigeria, and the Democratic Republic of the Congo, have integrated features into their DHIS2 platforms to identify and monitor ZD [29,30]. Globally, DHIS2 is widely adopted in more than 70 countries to strengthen health information systems and improve public health program management, including immunization services [31].

On the other hand, although CES 2019 provides comprehensive immunization data at the household level, its reliance on surveys and retrospective data collection introduces potential biases and scalability challenges. Recall bias and the inability to provide real-time data limit the accuracy and timeliness of the

XSL•FO
RenderX

information. Additionally, the resource-intensive nature of the survey makes it difficult to scale or repeat on a routine basis. Thus, CES is valuable but needs to be paired with real-time systems for more effective monitoring and interventions [13].

In addition, RCM is identified as a useful tool for filling immunization gaps as they appear because it allows real-time tracking of ZD children and the causes of nonvaccination. However, it has limited scope in particular regions, which reduces its generalizability at the national level [32]. In contrast, population-representative data from the BDHS 2017-2018 makes regional and temporal comparisons easier [33]. However, its reliance on immunization records, combined with the lack of data on nonvaccination, highlights a significant knowledge gap in understanding the causes of missed vaccinations. Although its methodology provides valuable insights, updates to the immunization tracking strategy are essential to address contemporary vaccination challenges. Understanding the in-depth reasons behind the high proportion of ZD children is crucial for effective policy formulation [34].

We found that HDSS provides valuable insights into population dynamics such as migration and their impact on vaccination coverage. However, its aggregated data limit its use for identifying ZD children at the household level. Integrating HDSS with more granular sources such as CES, BDHS, or RCM could improve the precision of immunization strategies. Enhancing HDSS to capture disaggregated data may further support efforts to reach underserved populations and strengthen vaccine equity. Moreover, although strong regulatory frameworks for VPD surveillance systems enhance vaccination distribution and safety monitoring, persistent systemic issues, such as cold chain limitations and geographic disparities, remain major barriers to achieving equitable coverage [35].

To strengthen monitoring systems, it will be crucial to integrate national platforms such as DHIS2 and BDHS with the localized insights offered by CES 2019 and RCM [13,15]. Although DHIS2 and BDHS capture broader immunization trends, CES and RCM provide critical microlevel data that can support the identification and understanding of ZD children [7]. However, achieving equitable vaccine coverage also demands overcoming systemic barriers such as data inaccessibility, workforce shortages, and geographic disparities [36,37]. Incorporating sociobehavioral insights into immunization tracking systems can further enhance their responsiveness to community-level barriers. A harmonized, multisource approach holds significant potential for reaching ZD populations, particularly in underserved and hard-to-reach areas [33].

Limitations

This study has several limitations. First, it did not follow a systematic review methodology such as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which may affect the comprehensiveness and reproducibility of the evidence synthesis process. Second, the qualitative analysis did not include direct quotations or attributed participant perspectives, which may have limited the depth of interpretation of stakeholder views. Finally, restricted access to certain immunization data sources limited the breadth of the analysis.

Strengths

Despite these limitations, this study provides a comprehensive assessment of Bangladesh's immunization data ecosystem by systematically mapping multiple routine, survey-based, and surveillance data systems. It combines a documentary review with stakeholder consultations, enabling triangulation of findings. The inclusion of diverse data sources allows a holistic evaluation of data quality, timeliness, and utility for ZD tracking and programmatic decision-making.

This study highlights clear opportunities to strengthen Bangladesh's immunization data ecosystem through improved integration and harmonization of existing systems. Linking routine platforms such as DHIS2 with survey and surveillance data could enhance data completeness, validation, and triangulation for ZD tracking. Expanding the use of unique identifiers and improving interoperability across systems would support better individual-level tracking. Incorporating contextual and sociobehavioral data into routine systems could further improve the targeting of underserved populations and inform more responsive immunization strategies.

Conclusions

This study highlights fragmentation across immunization data systems in Bangladesh, where multiple platforms capture different aspects of immunization, but none provide a complete, timely, and disaggregated picture of ZD children. Limitations include weak system integration, variable data quality (particularly denominator accuracy), and limited availability of subdistrict-level data, which constrain routine monitoring and equity-focused decision-making. DHIS2 shows the strongest potential for routine ZD tracking, while RCM offers more granular, context-specific insights and HDSS supports in-depth subpopulation analysis; CES provides periodic validation but requires better disaggregation and integration with routine systems. Strengthening interoperability across DHIS2, RCM, and CES; standardizing core indicators; improving data quality assurance; and enabling routine use of subdistrict-level data are essential steps to support a harmonized, multisource approach for improving ZD identification and immunization equity in Bangladesh.

Acknowledgments

The authors extend their gratitude to the collaborative partner, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), for their continuous support. The authors also gratefully acknowledge the cooperation and contributions of key stakeholders and participants from the World Health Organization (Bangladesh Country Office), United Nations Children's Fund

XSL•FO
RenderX

(Bangladesh Country Office), icddr,b, and the government of Bangladesh, particularly officials from the Ministry of Health and Family Welfare.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

This work was supported by Gavi, the Vaccine Alliance.

Authors' Contributions

SY conceptualized the research idea and led data curation. SY, PS, and ANJ performed data analysis and interpretation. SY and PS were responsible for writing the manuscript. SY, PS, ANJ, EO, MJU, NA, MWA, HD, and CM contributed to writing, reviewing, and revising the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Immunization coverage. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/immunization-coverage [accessed 2026-08-26]
2. A practical guide to identifying, addressing and tracking inequities in immunization. World Health Organization. 2024 Jul 12. URL: https://www.who.int/europe/publications/i/item/WHO-EURO-2024-9917-49689-74446 [accessed 2026-08-26]
3. Programme funding guidelines. Gavi. 2023 Apr. URL: https://www.gavi.org/sites/default/files/support/guidelines-2022/Gavi_Programme_Funding_Guidelines_ENG.pdf [accessed 2026-08-26]
4. Wigley A, Lorin J, Hogan D, Utazi CE, Hagedorn B, Dansereau E, et al. Estimates of the number and distribution of zero-dose and under-immunised children across remote-rural, urban, and conflict-affected settings in low and middle-income countries. PLOS Glob Public Health 2022 Oct 26;2(10):e0001126 [FREE Full text] [doi: 10.1371/journal.pgph.0001126] [Medline: 36962682]
5. Dong TQ, Wakefield J. Modeling and presentation of vaccination coverage estimates using data from household surveys. Vaccine 2021 Apr 28;39(18):2584-2594 [FREE Full text] [doi: 10.1016/j.vaccine.2021.03.007] [Medline: 33824039]
6. Zero-dose children and missed communities. Gavi. URL: https://www.gavi.org/our-alliance/strategy/phase-5-2021-2025/equity-goal/zero-dose-children-missed-communities [accessed 2026-08-26]
7. Country learning hub for immunization equity in Bangladesh: findings from rapid assessment. Gavi, the Vaccine Alliance. 2023. URL: https://zdlh.gavi.org/resources/country-learning-hub-immunization-equity-bangladesh-findings-rapid-assessment [accessed 2026-08-27]
8. Leaving no child behind: scaling up investment in immunization to protect against diseases. UNICEF. URL: https://www.unicef.org/bangladesh/media/5271/file/Immunization [accessed 2026-08-26]
9. Jannat Z, Das H, Ali MW, Alam N, Chowdhury ME, Sarker BK, et al. Identifying the zero-dose and under-immunized children in Bangladesh: approaches and experiences. PLoS One 2024 Oct 28;19(10):e0312171 [FREE Full text] [doi: 10.1371/journal.pone.0312171] [Medline: 39466797]
10. Diphtheria tetanus toxoid and pertussis (DTP) vaccination coverage. World Health Organization. URL: https://immunizationdata.who.int/global/wiise-detail-page/diphtheria-tetanus-toxoid-and-pertussis-(dtp)-vaccination-coverage [accessed 2026-08-26]
11. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008 Jul 21;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
12. E-tracker for maternal health and child immunization. Directorate General of Health Services Bangldesh. URL: https://epietracker.dghs.gov.bd/dhis-web-commons/security/login.action [accessed 2026-08-27]
13. Bangladesh EPI coverage evaluation survey 2019. Institute for Health Metrics and Evaluation. URL: https://ghdx.healthdata.org/record/bangladesh-epi-coverage-evaluation-survey-2019 [accessed 2026-08-27]
14. Immunization dashboard: Bangladesh. World Health Organization. URL: https://immunizationdata.who.int/dashboard/regions/south-east-asia-region/BGD [accessed 2026-08-26]
15. National Institute of Population Research and Training (NIPORT) and ICF. Bangladesh demographic and health survey 2022: final report. The DHS Program. 2022. URL: https://dhsprogram.com/publications/publication-FR386-DHS-Final-Reports.cfm [accessed 2026-08-27]
16. Health and Demographic Surveillance Systems (HDSS): Matlab. icddr,b. 2022. URL: http://dspace.icddrb.org/jspui/bitstream/123456789/11564/2/HDSS%20Annual%20Report%202020.pdf [accessed 2026-08-26]
17. Health and Demographic Surveillance Systems (HDSS). icddr,b. URL: https://www.icddrb.org/health-and-demographic-surveillance-systems [accessed 2026-08-27]

XSL•FO
RenderX

18. Effective vaccine management (EVM). World Health Organization. URL: https://www.who.int/southeastasia/activities/effective-vaccine-management-(evm) [accessed 2026-08-26]
19. National Guideline for AFP and Vaccine Preventable Diseases Surveillance Bangladesh. Dhaka, Bangladesh: Directorate General of Health Services; 2024.
20. Directorate General of Drug Administration (DGDA), Bangladesh. URL: https://www.dgdagov.info/ [accessed 2026-08-27]
21. Data quality self-assessment tool for national EPI programme. World Health Organization. URL: https://www.who.int/bangladesh/news/detail/30-06-2016-data-quality-self-assessment-tool-for-national-epi-programme [accessed 2026-08-26]
22. The immunization data quality self-assessment (DQS) tool. World Health Organization. 2005. URL: https://cdn.who.int/media/docs/default-source/immunization/immunization-coverage/dqs_tool.pdf?sfvrsn=41e43b5a_6 [accessed 2026-08-26]
23. The Geographical Information System (GIS) introduced in EPI program, Bangladesh. World Health Organization. 2023. URL: https://www.who.int/bangladesh/news/detail/07-08-2023-the-geographical-information-system-(gis)-introduced-in-epi-program-bangladesh [accessed 2026-08-26]
24. Working together: an integration resource guide for immunization services throughout the life course. World Health Organization. 2018. URL: https://iris.who.int/server/api/core/bitstreams/83c9d7d3-9bb7-487b-9195-79899f4430a1/content [accessed 2026-08-26]
25. Global progress report on WASH in health care facilities: fundamentals first. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240017542 [accessed 2026-08-26]
26. Nearly half a million children in Bangladesh miss full immunization, despite 81.6 % coverage: Gavi, UNICEF, and WHO urge immediate action for equity. UNICEF. 2025. URL: https://www.unicef.org/bangladesh/en/press-releases/nearly-half-million-children-bangladesh-miss-full-immunization [accessed 2026-08-26]
27. Zero-dose analysis card. Gavi, the Vaccine Alliance. URL: https://www.gavi.org/sites/default/files/support/Gavi_Zero-dose_AnalysisCard.pdf [accessed 2026-08-26]
28. Immunization Agenda 2030: progress towards global immunization goals and implementation of the Immunization Agenda 2030. World Health Organization. 2023. URL: https://apps.who.int/gb/ebwha/pdf_files/EB154/B154_9-en.pdf [accessed 2026-08-26]
29. Jean Baptiste AE, Wagai J, Hahné S, Adeniran A, Koko RI, de Vos S, et al. High-resolution geospatial mapping of zero-dose and underimmunized children following Nigeria's 2021 Multiple Indicator Cluster Survey/National Immunization Coverage Survey. J Infect Dis 2024 Jul 25;230(1):e131-e138. [doi: 10.1093/infdis/jiad476] [Medline: 39052714]
30. Using DHIS2 to reach zero-dose and under-immunized children in Mozambique. Gavi, the Vaccine Alliance. 2025. URL: https://www.gavi.org/vaccineswork/using-dhis2-reach-zero-dose-and-under-immunized-children-mozambique?utm_source=chatgpt.com [accessed 2026-08-26]
31. Dehnavieh R, Haghdoost A, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf Manag 2019 May;48(2):62-75. [doi: 10.1177/1833358318777713] [Medline: 29898604]
32. Estimating and monitoring yellow fever reactive campaign vaccination coverage: overview of survey and monitoring methods. World Health Organization. 2024. URL: https://www.who.int/publications/i/item/9789240090514 [accessed 2026-08-26]
33. Bangladesh demographic and health survey 2017–18: key indicators. National Institute of Population Research and Training. 2019 Nov. URL: https://dhsprogram.com/pubs/pdf/PR104/PR104.pdf [accessed 2026-08-26]
34. Biks GA, Shiferie F, Tsegaye DA, Asefa W, Alemayehu L, Wondie T, et al. In-depth reasons for the high proportion of zero-dose children in underserved populations of Ethiopia: results from a qualitative study. Vaccine X 2024 Feb 1;16:100454 [FREE Full text] [doi: 10.1016/j.jvacx.2024.100454] [Medline: 38327767]
35. Global strategy for comprehensive Vaccine-Preventable Disease (VPD) surveillance. World Health Organization. 2020. URL: https://www.who.int/publications/m/item/global-strategy-for-comprehensive-vaccine-preventable-disease-(vpd)-surveillance [accessed 2026-08-26]
36. Ronveaux O, Rickert D, Hadler S, Groom H, Lloyd J, Bchir A, et al. The immunization data quality audit: verifying the quality and consistency of immunization monitoring systems. Bull World Health Organ 2005 Jul;83(7):503-510 [FREE Full text] [Medline: 16175824]
37. National Academies of Sciences, Engineering, and Medicine. The Critical Public Health Value of Vaccines: Tackling Issues of Access and Hesitancy: Proceedings of a Workshop. Washington, DC: The National Academies Press; 2021.

Abbreviations

AEFI: adverse event following immunization
BDHS: Bangladesh Demographic and Health Survey
CES: Coverage Evaluation Survey
CLH: Country Learning Hub
DHIS2: District Health Information Software 2
DTP1: first dose of the diphtheria, tetanus, and pertussis

XSL•FO
RenderX

EPI: Expanded Programme on Immunization
HDSS: Health and Demographic Surveillance System
icddr,b: International Centre for Diarrhoeal Disease Research, Bangladesh
LMIC: low- and middle-income country
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCM: Rapid Convenience Monitoring
SIMO: Surveillance and Immunization Medical Officer
UNICEF: United Nations Children's Fund
VPD: vaccine-preventable disease
WHO: World Health Organization
ZD: zero dose

Edited by E Mensah; submitted 04.Dec.2025; peer-reviewed by M Das, K Aninkora, A Bethanabatla; comments to author 29.May.2026; revised version received 16.Jun.2026; accepted 20.Jun.2026; published 29.Sep.2026.

Please cite as:
Yunus S, Singh P, John AN, Oliveras E, Uddin MJ, Alam N, Ali MW, Das H, Morgan C
Using Immunization Data Systems to Track Zero-Dose Children in Bangladesh: Mixed Methods Exploratory Study
Online J Public Health Inform 2026;18:e88951
URL: https://ojphi.jmir.org/2026/1/e88951
doi:10.2196/88951
PMID:

©Samiha Yunus, Prem Singh, Alan Noble John, Elizabeth Oliveras, Md Jasim Uddin, Nurul Alam, Md Wazed Ali, Hemel Das, Chris Morgan. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 29.Sep.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigation of Community Behaviors, Socioeconomic Factors, and Breakthrough COVID-19 Infections Among Vaccinated Individuals: Cross-Sectional Study

Matthew J McDonald, MS, DScPAS, PA-C; Bathri Narayan Vajravelu, MBBS, MPH, PhD

Department of Physician Assistant Studies, Massachusetts College of Pharmacy and Health Sciences, Boston, MA, United States

Corresponding Author:
Bathri Narayan Vajravelu, MBBS, MPH, PhD
Department of Physician Assistant Studies, Massachusetts College of Pharmacy and Health Sciences, Boston, MA, United States

Abstract

Background: Despite widespread COVID-19 vaccination, breakthrough infections remain a public health concern, with transmission risks potentially linked to community behaviors and age-specific preventive practices. While mask-wearing and social distancing are well-established mitigation strategies, their adoption patterns across age groups, particularly among vaccinated individuals, are poorly understood.

Objective: This study focuses on understanding breakthrough infections among vaccinated individuals, high-risk behaviors, and socioeconomic determinants of COVID-19 susceptibility to guide effective public health interventions.

Methods: A 31-question voluntary survey was distributed using convenience sampling through the Qualtrics survey platform. All survey respondents reported receiving at least the primary vaccination against COVID-19 infection, and all survey responses were recorded between January 6, 2022, and September 26, 2022. Logistic regression analysis was used to estimate the odds ratio to measure the association between testing positive for COVID-19 and the different activities.

Results: Among the vaccinated individuals, those who tested positive were 11.103 times more likely to engage in going to a restaurant or bar compared to those who tested negative (P=.01). There was a significant difference in practicing social distancing and mask-wearing between the different age groups (P=.02), with 100% (10/10) of the participants older than 70 years practicing it, followed by 96.8% (118/122) of the 18 to 29 year olds. The study found lower infection rates in the same age groups compared to the other age groups. Moreover, the 18 to 29 years age group demonstrated notable associations with practicing social distancing and mask-wearing in various settings.

Conclusions: Compliance with social distancing and mask-wearing was higher among older and younger participants, and noncompliance with social distancing and mask-wearing was associated with a higher positivity rate. Activities such as going to a restaurant or bar were significantly associated with testing positive for COVID-19 among vaccinated individuals. These results provide valuable information to individuals, health care providers, and public health experts regarding the types of behaviors and community settings that are associated with COVID-19 infection and help enhance our understanding of the types of settings in which social distancing and masking may be beneficial or not necessary. This knowledge can also help local health departments develop tailored public health guidance based on the behaviors of individuals and the types of community settings in their localities.

(Online J Public Health Inform 2026;18:e76679) doi:10.2196/76679

KEYWORDS

high-risk behaviors; vaccinated individuals; socioeconomic factors; breakthrough infections; COVID-19; social distancing

Introduction

Since late 2019, the emergence of coronavirus disease 2019, commonly known as COVID-19, has disrupted much of society and resulted in over 1.1 million reported deaths in the United States [1]. Early in the pandemic, the Centers for Disease Control and Prevention (CDC), local public health departments, and other public health authorities sought to mitigate transmission and potential loss through various public health measures, which included public masking requirements, social distancing, lockdowns, and school and business closures.

The transmission of COVID-19 is variable and dependent on many factors, including the type of contact with an individual who is infected, duration of exposure, host immune factors, and ventilation [2]. The dominant route of transmission of COVID-19 is respiratory, and it is known that COVID-19 spreads effectively in enclosed and crowded settings [3]. Masking and social distancing have been shown to reduce transmission of COVID-19 in many settings [4-6].

XSL•FO
RenderX

The Omicron variants, known for increased transmissibility and immune evasion from both natural and vaccine-induced immunity, were discovered in 2022 and became the dominant variant across the world in 2023 [7]. The CDC tracked new cases of COVID-19 until early 2023. Guidance on personal protective measures had previously been based on community levels of circulating COVID-19 infections. However, in 2023, the CDC changed its data collection and reporting, focusing on the weekly number of hospitalizations and deaths rather than new cases of infection, as well as monitoring variants based on data from a limited number of areas in the United States [8]. With the dominance of Omicron, COVID-19 vaccine formulations were updated from previously protecting against the XBB sublineage strains to providing protection against the Omicron JN.1 lineage [9]. On June 27, 2024, the CDC's Advisory Committee on Immunization Practices recommended vaccination with a Food and Drug Administration–approved or authorized vaccine for all persons aged 6 months or older [9]. Current Food and Drug Administration–approved vaccinations for COVID-19 prevention include vaccines by Moderna, Pfizer-BioNTech, and Novavax. Prior COVID-19 infection as well as COVID-19 vaccination have been shown to reduce the severity and transmission of COVID-19 infection, and this may impact the accuracy of CDC COVID-19 data collection [10].

Many of the stringent public health restrictions implemented early in the pandemic have been replaced by less strict, practical measures that place the responsibility of social distancing and masking when sick on individuals in conjunction with local transmission rates [2]. In March 2024, the CDC released updated recommendations for individuals to prevent the spread of upper respiratory viruses from common respiratory viruses, including COVID-19, influenza, and respiratory syncytial virus, intending to simplify public health recommendations rather than have separate pathogen-specific guidelines [9]. The guidelines allow individuals who have confirmed or suspected COVID-19 infection to resume normal activities if their symptoms are improving and they have been afebrile (without antipyretics) for at least 24 hours. They are encouraged to take additional prevention strategies for the next 5 days, such as masking, enhancing hygiene practices, and social distancing [9]. With newer guidelines allowing individuals the ability to re-enter the community while still symptomatic and potentially infectious with upper respiratory viruses, including COVID-19, research pertaining to specific community settings and behaviors that increase one's risk for infection is necessary to help individuals reduce their own risk of infection [11].

With the availability of COVID-19 vaccines, understanding the types of community behaviors and settings that are associated with a high risk of COVID-19 transmission in vaccinated individuals is an important area of research because individuals can make more informed decisions about the types of behaviors and settings they are willing to engage in based on the risk of infection even after vaccination. Individuals will be better informed about the potential risk of infection, allowing them to determine whether personal protective measures to reduce the risk of infection, such as social distancing or masking, are more appropriate in specific settings. For policymakers and public health experts, this study will provide information about the behaviors and settings that are associated with a risk of COVID-19 infection, allowing them to develop tailored infection prevention strategies and guidance for community members pertaining to COVID-19 and other respiratory viral infections.

This study aims to identify community behaviors, settings, and close contact exposures of vaccinated individuals that are associated with testing positive for COVID-19 despite vaccination status. Additionally, this study sought to evaluate any socioeconomic factors associated with COVID-19 infection, including demographics, household income, and type of residence. The authors' hypothesis is that certain community behaviors that have the potential to be crowded and have poor ventilation, such as indoor restaurants, bars, cafés, places of worship, fitness centers, entertainment venues, and using public transportation, will be associated with a higher breakthrough risk of COVID-19 infection.

Methods

Ethical Considerations

The institutional review board of Massachusetts College of Pharmacy and Health Sciences (MCPHS) in Boston, MA, reviewed and approved this cross-sectional research study on November 19, 2021 (IRB-2021-2022-39). This study was conducted in accordance with the Belmont Report ethical principles of respect for persons, beneficence, and justice, and in compliance with the Federal Policy for the Protection of Human Subjects (45 CFR 46, The Common Rule). The study also adhered to the principles of the Declaration of Helsinki and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Multimedia appendices 1). Informed consent was obtained electronically from all participants prior to their participation via the Qualtrics online survey platform. Participants were informed of the study's purpose, voluntary nature, and their right to withdraw at any time without consequence. No compensation was provided for participation. To protect participant privacy and confidentiality, no personally identifiable information was collected at any point during the study. All survey responses were collected anonymously through Qualtrics, which was configured to disable the collection of IP addresses and other identifying metadata. Data were stored securely on password-protected institutional servers, accessible only to the research team. All data were analyzed in aggregate and reported in a manner that prevents the identification of individual participants.

Study Design and Data Collection

An online survey was created using Qualtrics, and all survey responses were recorded between January 6, 2022, and September 26, 2022. The survey items related to sociodemographic characteristics, community behaviors, and personal preventive measures were adapted from previously validated survey instruments [12-14]. The survey was pilot-tested among a small group of faculty members and students at MCPHS to confirm the clarity and validity of the tool before distribution. The reliability of the survey was assessed using the Cronbach α coefficient, which was calculated to be 0.87, indicating good internal consistency among the survey items. Convenience sampling was used to recruit

XSL•FO
RenderX

respondents for the survey and included email, social media postings, and in-person recruitment. No formal sample size calculation or power analysis was performed before the study since this was an exploratory cross-sectional study using convenience sampling. The final sample size achieved was adequate for the regression analysis for the number of outcome events and the predictor variables included in the model. Approximately 823 individuals were recruited by email, and 5 social media postings were disseminated using Blackboard and Facebook. The recruitment emails and social media postings included information about the study and the survey link to participate. The social media postings targeted students and faculty at MCPHS. Recruitment emails targeted department heads and faculty members from several United States medical universities and physician assistant programs. The publicly available email addresses of department heads and faculty members from several United States medical universities and physician assistant programs were gathered from their respective university websites. Survey links could then be disseminated to other eligible individuals at their discretion. The survey consisted of a range of 22 to 31 questions. Survey items on frequency of activities (eg, visiting restaurants, coffee shops, gyms, and places of worship) used a 5-point Likert scale (Never, Rarely, Sometimes, Often, Always). For regression analyses, these responses were dichotomized into "Never/Rarely/Sometimes" versus "Most of the time/Always." This approach was chosen to improve model stability and ensure sufficient sample size within each category. Inclusion criteria included being at least 18 years old, and each respondent reported receiving at least the primary vaccination against COVID-19 infection. Primary vaccination included the original vaccine regimens in the United States: 2 doses of the original mRNA vaccines by Pfizer-BioNTech or Moderna, or 1 dose of the Johnson & Johnson vaccine. Respondents were included if they reported receiving a vaccine regimen approved in a non-US country, but the type of vaccine was not collected. A survey answer option of "Other" was included in the question where respondents were asked about receiving a vaccine that was not specifically listed. COVID-19 positivity was self-reported and based on a positive COVID-19 PCR or rapid antigen test anytime during the study period. The survey was closed, and no further responses were recorded once a second booster shot was authorized, as our survey did not include questions regarding a second booster, which could confound the study's results.

Statistical Analysis

Logistic regression analysis was used to estimate the odds of COVID-19 infection in relation to various risk factors identified within the survey dataset. The primary outcome variable was self-reported COVID-19 positivity (yes/no), and the primary exposures were self-reported frequency of participation in community activities (eg, shopping, dining at restaurants, bars, attending gyms, salons, air travel, public transportation, indoor entertainment, places of worship, and social gatherings), masking, and social distancing practices. Predictor variables were sociodemographic variables such as age, sex, and occupational exposure. Potential confounders were age and occupational exposure. There were no identified effect modifiers. Odds ratios with 95% CIs were reported for each analysis. Both univariable and multivariable logistic regression models were conducted. In the multivariable analyses, sociodemographic factors and behavioral risk variables were entered simultaneously to assess their independent associations with COVID-19 positivity. To minimize overfitting, we performed separate multivariable models focusing on conceptually grouped predictors (eg, community activities, social distancing behaviors, sociodemographic variables). Sociodemographic factors were not included in the same model as activity variables because of collinearity concerns and sample size limitations. This analytic strategy allowed a clearer interpretation of specific domains of risk. Respondents who did not complete the entire survey were removed from the data analysis. A P value of less than .05 was established for statistical significance. IBM SPSS Statistics version 29.0 (IBM Corp) was used for statistical analysis.

Results

The primary goal of this investigation is to unveil the community behaviors, settings, and close contact scenarios correlated with positive COVID-19 test results among vaccinated individuals. Additionally, the study explores the socioeconomic factors related to COVID-19 infection, including demographics, household income, and type of residence. A total of 297 participants completed the survey, of whom 25 were excluded because of missing data. Among the remaining 272 participants who answered the entire survey, 208 (76.5%) were female, 62 (22.8%) were male, and 2 (0.8%) identified as either nonbinary, third gender, or preferred not to say. The precise response rate is unclear because the researchers were unable to determine how many participants received the survey as a result of email or social media recruitment. The majority of respondents to the survey fell within the 18 to 29 years age group, constituting 44.9% (122/272) of participants, a demographic typically representative of college students. Furthermore, a substantial proportion of the survey participants identified as White, accounting for 83.8% (228/272), as shown in Table 1.

XSL•FO
RenderX

Table . Demographic characteristics of vaccinated adult participants in a cross-sectional online survey assessing breakthrough COVID-19 infections in the United States between January 2022 and September 2022.

Characteristic	Participants, n (%)
Gender	
Male	62 (22.8)
Female	208 (76.5)
Nonbinary/third gender	1 (0.4)
Prefer not to say	1 (0.4)
Age group (y)	
18 - 29	122 (44.9)
30 - 39	39 (14.3)
40 - 49	39 (14.3)
50 - 59	30 (11)
60 - 69	31 (11.4)
>70	10 (3.7)
Race or ethnicity	
Asian Indian	12 (4.4)
Black or African American	7 (2.6)
Chinese	5 (1.8)
Filipino	1 (0.4)
Hispanic or Latino	9 (3.3)
Multiracial or biracial	3 (1.1)
Native American or Alaskan Native	2 (0.7)
Vietnamese	1 (0.4)
White	228 (83.8)
Other race not listed	2 (0.7)

A logistic regression was used to assess the impact of community activities on the likelihood of COVID-19 infection in vaccinated individuals. The model exhibited statistical significance (χ^2_{11}=25.048; P<.001), indicating its ability to elucidate 33% (Nagelkerke R^2) of the variance in COVID-19 among vaccinated individuals and correctly classify 79.2% (215/272) of cases. The analysis provided insights into the influence of specific activities (see Table 2) on COVID-19 infection in vaccinated individuals. Notably, those who tested positive were 11.103 times more likely (P=.01) to engage in going to a restaurant or bar compared with those who tested negative (Table 2). These results suggest a significant association between community activities and COVID-19 infection in vaccinated individuals, emphasizing the importance of specific behaviors in influencing infection outcomes.

XSL•FO
RenderX

Table . Association between community activities and self-reported COVID-19 positivity among vaccinated adults who participated in the online cross-sectional survey in the United States between January 2022 and September 2022.

Activity	OR[a] (95% CI)	*P* value
Shopping for items	0.99 (0.27 - 3.65)	.99
Place of worship	0.24 (0.07 - 0.91)	.04
Restaurant or bar	11.10 (1.80 - 68.57)	.01
Coffee shop	0.77 (0.24 - 2.44)	.66
Public transport	0.24 (0.06 - 1.02)	.05
Airplane travel	0.49 (0.14 - 1.73)	.27
Office setting	3.08 (0.60 - 15.86)	.18
Gym or fitness	0.14 (0.02 - 0.93)	.04
Salon or barber	1.56 (0.47 - 5.17)	.47
Vehicle travel	2.39 (0.72 - 7.98)	.16
Indoor entertainment	0.35 (0.10 - 1.26)	.11

[a]OR: odds ratio.

Among the different age groups who responded to the survey, we found that infection rates were the lowest in the more than 70 years age group (30%; n=3) and the 18 to 29 years age group (40.7%; n=49) compared to the other age groups, although the difference did not achieve statistical significance. Then, we sought to determine whether there was any difference in the positivity rate associated with a specific activity between respondents who reported social distancing or wearing their masks most of the time and those who did not do so most of the time. The initial model, which included no predictors (null model), displayed a classification table with an overall correct classification rate of 59.9% (163/272), indicating suboptimal performance in predicting COVID-19 positivity based on the specified variables.

After incorporating predictors from the question that asked about "social distancing and mask-wearing behaviors in specific settings," the model showed statistically insignificant results (χ^2_{11}=17.612; P >.05). The model elucidated only 8.5% (Nagelkerke R^2) of the variance in COVID-19 positivity among vaccinated individuals and accurately classified 60.7% (165/272) of cases. A detailed examination of the results (Table 3) indicated that none of the predictors achieved statistical significance. For example, participants who tested positive were 3.001 times more likely to go to an office setting compared to those who did not test positive, but this result did not reach statistical significance (P=.09).

Table . Association between masking and social distancing and self-reported COVID-19 positivity among vaccinated adults who participated in the online cross-sectional survey in the United States between January and September 2022.

Activity	OR[a] (95% CI)	*P* value
Shop for items	0.54 (0.19 - 1.57)	.26
Place of worship	1.29 (0.42 - 3.94)	.65
Restaurant or bar	1.71 (0.48 - 6.11)	.41
Coffee shop	1.38 (0.39 - 4.88)	.62
Public transport	1.71 (0.46 - 6.29)	.42
Airplane travel	1.77 (0.52 - 6.04)	.36
Office setting	3.00 (0.85 - 10.65)	.09
Gym or fitness	1.09 (0.36 - 3.30)	.88
Salon or barber	0.87 (0.27 - 2.78)	.82
Vehicle travel	1.46 (0.45 - 4.72)	.53
Indoor entertainment	0.25 (0.07 - 0.87)	.03

[a]OR: odds ratio.

In summary, the logistic regression model did not provide substantial evidence of a significant difference in COVID-19 positivity rates based on specific activities among respondents reporting varying levels of social distancing and mask-wearing. The overall model did not achieve statistical significance, and the predictors included did not significantly contribute to explaining the variability in COVID-19 positivity.

XSL•FO
RenderX

We then tested whether any of the following demographic variables were associated with a higher COVID-19 positivity rate, including household income, type of residence, education level, race, and number of individuals living in a household. The logistic regression aimed to investigate whether specific demographic variables were correlated with a higher COVID-19 positivity rate among participants. The initial model, representing the null model with no predictors, exhibited a classification table with an overall correct classification rate of 59.9% (163/272), indicating suboptimal performance in predicting COVID-19 positivity based on the specified variables. Subsequent inclusion of demographic variables also yielded statistically insignificant findings (χ^2_{22}=20.573; P>.05). The model elucidated 10.6% (Nagelkerke R^2) of the variance in COVID-19 positivity among participants and accurately classified 61% (166/272) of cases. The analysis of the demographic variables (Table 4) revealed that none of the predictors achieved statistical significance. This suggests that none of the demographic variables were independently associated with a higher COVID-19 positivity rate.

Table . Association between sociodemographic factors and self-reported COVID-19 positivity among vaccinated adults who participated in the online cross-sectional survey in the United States between January 2022 and September 2022.

Variable	OR[a] (95% CI)	*P* value
Annual income (US $)		
Under 25,000	1.14 (0.42 - 3.09)	.79
25,000 - 49,999	0.43 (0.11 - 1.68)	.22
50,000 - 74,999	1.33 (0.41 - 4.36)	.64
75,000 - 99,999	0.66 (0.17 - 2.64)	.56
100,000 - 149,999	1.22 (0.35 - 4.32)	.75
150,000 - 199,999	1.09 (0.41 - 2.92)	.86
Over $200,000	0.64 (0.19 - 2.19)	.48
Residence type		
Single-family house	1.15 (0.21 - 6.31)	.87
Two-family home or duplex	0.92 (0.49 - 1.73)	.79
Apartment or Condo	0.00 (0.00-∞)	>.99
Education		
High school	2.28 (0.00-∞)	>.99
Some college—no degree	2,082,843,782.47 (0.00-∞)	>.99
Associate degree	1.18 (0.26 - 5.40)	.84
Bachelor's degree	1.81 (0.39 - 8.39)	.45
Graduate degree	2.63 (0.21 - 32.61)	.45
Ethnicity		
White	0.00 (0.00-∞)	>.99
Black/African American	0.31 (0.06 - 1.63)	.17
Hispanic/Latino	1.37 (0.25 - 7.39)	.71
Native American/Alaskan Native	1.03 (0.15 - 7.16)	.98
Asian Indian	0.00 (0.00-∞)	>.99
Multiracial/Biracial	0.37 (0.07 - 1.90)	.23
Chinese	0.00 (0.00-∞)	>.99
Filipino	1.12 (0.04 - 28.39)	.95
Vietnamese	1.33 (0.08 - 22.78)	.85
Other race not listed	0.00 (0.00-∞)	>.99

[a]OR: odds ratio.

In summary, the logistic regression model revealed no statistically significant associations between demographic variables (eg, income, education, and race) and COVID-19 positivity in our vaccinated cohort. While this suggests that socioeconomic factors alone may not independently predict breakthrough infections in this population, it underscores the

XSL•FO
RenderX

potential dominance of behavioral or environmental determinants (eg, activity-related risks, as identified in our other analyses).

We then investigated whether living with someone who worked in health care was associated with a higher positivity rate. Contrary to expectations, our analysis found no significant association between living with a health care worker and COVID-19 positivity rates among vaccinated individuals (Table 5). This suggests that, in our cohort, household exposure to health care personnel, who likely had frequent occupational exposure to SARS-CoV-2, did not translate into a higher infection risk for vaccinated cohabitants. These findings may reflect the success of infection control measures in health care settings, the protective effect of vaccination in household contexts, or potential compensatory behaviors (eg, stricter hygiene practices) in these households. While reassuring, these results should be interpreted in context: the study period (2022) coincided with widespread vaccine availability and improved workplace protections, which may have mitigated risks that were more prominent earlier in the pandemic.

Table . Association between household exposure to health care workers and self-reported COVID-19 positivity among vaccinated adults who participated in the online cross-sectional survey in the United States between January 2022 and September 2022.

Household health care workers	OR[a] (95% CI)	*P* value
None	0.00 (0.00-∞)	>.99
One	1.15 (0.65 - 2.03)	.64
More than one	0.91 (0.38 - 2.19)	.83

[a]OR: odds ratio.

We also tested whether there was an association between a specific vaccine (Moderna, Pfizer, or Johnson & Johnson) and a higher or lower hospitalization rate among respondents. The logistic regression model did not yield evidence of a significant association between the type of vaccine and the hospitalization rate among participants. The overall model was statistically insignificant, and the predictor included in the model did not substantially contribute to explaining the variability in hospitalization rates among the participants (Table 6). This indicates that all 3 vaccines are equally effective in reducing hospitalizations due to COVID-19.

Table . Association between type of COVID-19 vaccine and hospitalization rate among vaccinated adults who participated in the online cross-sectional survey in the United States between January 2022 and September 2022.

Vaccine	OR[a] (95% CI)	*P* value
Moderna	0.98 (0.27 - 3.61)	.98
Pfizer-BioNTech	2.21 (0.60 - 8.12)	.23
Johnson & Johnson	0.87 (0.51 - 1.47)	.60

[a]OR: odds ratio.

We then analyzed the relationship between distinct age groups and their adherence to social distancing and mask-wearing practices in various scenarios. We identified a statistically significant difference among the different age groups in practicing social distancing and mask-wearing practices (*P*=.02). The highest prevalence was found in the older than 70 years age group, where 100% of the respondents reported social distancing practices, followed by the 18 to 29 years age group (96.8%). The results reveal significant associations, particularly within the 18 to 29 age group, consistently exhibiting a higher likelihood of practicing these preventive measures across diverse settings. When it comes to shopping for items, individuals aged 18 to 29 consistently showed a significant association with “Always or Most of the time,” engaging in social distancing and mask-wearing. Similarly, this age group displayed significant associations in scenarios involving people gathering indoors, whether with more than 10 people or with 10 people or fewer.

Moreover, the 18 to 29 age group demonstrated notable associations with practicing social distancing and mask-wearing in various settings, including attending indoor church or religious gatherings, going to restaurants or bars, visiting coffee shops, using public transportation, traveling via airplane, going to an office setting (excluding health care purposes), going to a gym or fitness center, and visiting a salon or barber. The consistent pattern emphasizes the inclination of the 18 to 29 age group to adhere to social distancing and mask-wearing guidelines across a spectrum of activities. The statistical significance of these associations, as indicated by the *P* value, underscores the robustness of these observed patterns. These insights can inform targeted public health interventions and communication strategies, recognizing the variations in behavior across different age demographics (Table 7).

XSL•FO
RenderX

Table . Association between age groups and their adherence to social distancing and mask-wearing practices in various community settings among vaccinated adults who participated in the online cross-sectional survey in the United States between January 2022 and September 2022.

Activity and age group (y)	Never or rarely or sometimes	Always or most of the time	*P* value
Shop for items (groceries, prescriptions, home goods, and clothing)			<.001
>70	1	9	
18 - 29	24	98	
30-39	2	37	
40 - 49	0	39	
50 - 59	0	30	
60 - 69	1	30	
Have people visit you inside your home or go inside someone else's home where there were more than 10 people			<.001
>70	2	8	
18 - 29	43	79	
30-39	2	37	
40 - 49	0	39	
50 - 59	1	29	
60 - 69	1	30	
Have people visit you inside your home or go inside someone else's home where there were 10 people or less			<.001
>70	1	9	
18 - 29	35	87	
30-39	2	37	
40 - 49	1	38	
50 - 59	2	28	
60 - 69	2	29	
Go to an indoor church or a religious gathering or place of worship			<.001
>70	0	10	
18 - 29	33	89	
30-39	2	37	
40 - 49	0	39	
50 - 59	2	28	
60 - 69	2	29	
Go to a restaurant or bar (dine-in, any area designated by the restaurant including patio seating)			<.001
>70	2	8	
18 - 29	29	93	
30-39	1	38	
40 - 49	0	39	
50 - 59	1	29	
60 - 69	1	30	
Go to a coffee shop			<.001
>70	1	9	
18 - 29	28	94	
30-39	1	38	
40 - 49	0	39	
50 - 59	0	30	

XSL•FO
RenderX

Activity and age group (y)	Never or rarely or sometimes	Always or most of the time	*P* value
60 - 69	1	30	
Use public transportation (bus, subway, streetcar, and train)			<.001
>70	1	9	
18 - 29	26	96	
30-39	1	38	
40 - 49	0	39	
50 - 59	1	29	
60 - 69	2	29	
Travel via airplane			<.001
>70	2	8	
18 - 29	42	80	
30-39	2	37	
40 - 49	0	39	
50 - 59	2	28	
60 - 69	1	30	
Go to an office setting (other than for health care purposes)			<.001
>70	0	10	
18 - 29	36	86	
30-39	2	37	
40 - 49	2	37	
50 - 59	1	29	
60 - 69	1	30	
Go to a gym or fitness center			<.001
>70	0	10	
18 - 29	40	82	
30-39	1	38	
40 - 49	1	38	
50 - 59	2	28	
60 - 69	2	29	
Go to a salon or barber (eg, hair salon and nail salon)			<.001
>70	1	9	
18 - 29	43	79	
30-39	2	37	
40 - 49	0	39	
50 - 59	0	30	
60 - 69	2	29	

Discussion

Overview

Determining the type of community activities and settings that are associated with COVID-19 in vaccinated individuals is important for individuals to assess the risk of the activities they are willing to engage in. Identifying the behaviors and community settings that are associated with COVID-19 infection will also help enhance our understanding of the types of settings in which social distancing and masking may be more efficacious. This knowledge can help local health departments develop tailored public health guidance based on the behaviors of

XSL•FO
RenderX

individuals and the types of community settings in their localities.

Principal Results

One of the primary goals of this study was to reveal community behaviors linked to positive COVID-19 test results among vaccinated individuals. The present study demonstrates that vaccinated individuals were more likely to get infected with the virus while visiting a restaurant or a bar, emphasizing the importance of specific behaviors in shaping infection outcomes. This explains the infection rate among vaccinated individuals and aligns with the results reported by Zhang et al [15,16], who found higher infection rates among people visiting a restaurant. This might be a result of inadequate ventilation, which could lead to a rise in the concentration of airborne SARS-CoV-2. The current research revealed that individuals from the youngest (18 - 29 y) and oldest (>70 y) demographics were notably diligent about practicing social distancing behavior and mask-wearing, potentially linked to elevated levels of COVID-19-related anxiety among this cohort. This might also stem from the fact that individuals in the youngest age group are highly socially connected and actively engaged in extensive community interactions [14].

We sought to examine the difference in positivity rates related to specific activities between respondents who reported adhering to social distancing and wearing masks most of the time and those who did not. Nevertheless, our analysis did not provide strong evidence of a significant difference in COVID-19 positivity rates based on specific activities among respondents with varying levels of social distancing and mask-wearing. The overall model did not achieve statistical significance, and the predictors included did not substantially contribute to explaining the variability in COVID-19 positivity. Nevertheless, the present study revealed lower infection rates among both the oldest and youngest age groups, which coincided with their high adherence to social distancing measures. These findings are well corroborated by a recent study that found that each 1 unit increase in social distancing was associated with a 26% reduced risk of COVID-19 incidence and a 31% reduced risk of COVID-19 mortality at the county level [13]. Additionally, another study observed that COVID-19 epidemic case growth rates declined by approximately 1% per day, beginning 4 days after the implementation of state-wide social distancing measures [12]. These findings contribute to the existing body of evidence by estimating the impact of social distancing in the community on individual-level outcomes.

Comparison With Prior Work

Numerous studies have consistently indicated a connection between face mask usage and a decreased risk of COVID-19 at the population level. Specifically, 3 prior studies examining the impact of self-reported face mask use on the risk of COVID-19 reported odds ratios ranging from 0.21 to 0.30, which align with our finding of a 0.36 hazard ratio for always using masks [17-19]. While the protective effect of face masks in reducing transmission to others is well established, our findings suggest that mask use may also reduce the wearer's exposure to viral load, potentially contributing to decreased infection risk, as previouslyproposed [20].

This study investigated whether demographic variables were associated with a higher COVID-19 positivity rate. However, the logistic regression model did not provide significant evidence of an association between demographic variables (household income, type of residence, education level, race, and number of individuals living in a household) and COVID-19 positivity. The overall model was statistically insignificant, and the predictors included did not substantially contribute to explaining the variability in COVID-19 positivity among participants. Our study results also did not provide evidence of a significant association between living with someone in health care and the COVID-19 positivity rate among participants.

This study also explored whether there is an association between a specific vaccine (Moderna, Pfizer, or Johnson & Johnson) and hospitalization rates. However, we did not find evidence of a significant association between the type of vaccine and hospitalization rates among participants. This indicates that all 3 vaccines are similarly effective in preventing severe outcomes that require hospitalization. It suggests that public health benefits can be achieved with any of the available vaccines and supports their continued use in efforts to reduce the burden of severe illness.

The unprecedented speed and scale of data accumulation on breakthrough infections and related topics have not answered several important questions. For instance, though there is evidence that vaccines against SARS-CoV-2 reduce transmission in households [21] and communities [22], achieving sustained high levels of herd immunity against SARS-CoV-2 infection through vaccination is questioned. This skepticism arises from the mucosal nature of the infection without an obligate stage of dissemination through lymph or blood. Even with high vaccine coverage, a combination of waning immunity and antigenic variation may create enough susceptibility in the population to maintain endemic transmission of SARS-CoV-2, similar to the 4 other coronaviruses circulating in the human population [23]. However, it is unlikely that this situation will result in the same level of disruption seen in the first 2 years of the COVID-19 pandemic. Pandemics are rare events where virtually all humans lack exposure to a novel pathogen, putting them at risk for severe disease and transmission, especially older adult individuals and those with certain comorbidities. Similar to the influenza virus [24], or even more so with human coronaviruses, this pandemic pattern may gradually transition into a pattern of milder disease. Virtually everyone will experience multiple exposures through one or more vaccine doses and/or one or more exposures to viral (possibly breakthrough) infection [25]. In this view, the role of vaccines is not to provide durable herd immunity, as with measles or smallpox, but to prevent severe outcomes during the transition to endemicity.

Limitations

It is important to note the limitations of the study. This study used convenience sampling, primarily recruiting participants affiliated with academic institutions (eg, students and faculty), which allowed the researchers to focus recruitment on a population that had a known high vaccination rate but may limit the generalizability of the findings. The overrepresentation of younger, female, and White individuals in our cohort (Table 1)

XSL•FO
RenderX

could introduce selection bias, as behavioral responses to COVID-19 (eg, mask-wearing and dining out) may systematically differ in older or more diverse populations. Additionally, the lack of diversity in gender, race or ethnicity, age, geographical, and socioeconomic representation among the participants is another limitation that could impact the generalizability of the findings. For instance, our observed high adherence to social distancing among those aged more than 70 years (100%) should be interpreted cautiously, given the small sample size (n=10) in this subgroup. Many participants were affiliated with medical or other health care-related universities, so the participants may have been at higher risk for exposure to COVID-19 infection through occupational or study-related exposures rather than exposures in the community. Participants may have been more knowledgeable about public health guidance and COVID-19 mitigation strategies and more likely to follow masking and social distancing guidelines than the broader population. Since the study was conducted using an anonymous survey, we could not verify responses or the vaccination status of each respondent. Recall bias was mitigated by using validated surveys as a resource and by comparing the behaviors and community settings of respondents who reported testing positive for COVID-19 infection and those who had never tested positive for COVID-19. However, it is possible that there could be some inaccuracies in the responses from both cohorts of respondents. Our analytic approach used separate multivariable models for behavioral and socio-demographic factors rather than a single combined model. This was chosen to reduce collinearity and preserve model stability given the modest sample size. However, this limits our ability to draw conclusions about the independent effects of behavior after adjusting for demographics. Public health measures, such as public masking, social distancing, and reduced maximum occupancy limits for social gathering settings, were variable across the United States, and the degree of adherence to those recommendations was also variable, which could confound the study results regarding which settings were more or less likely to be associated with COVID-19 infection in vaccinated individuals. However, this heterogeneity was distributed evenly among both cohorts of respondents. Future studies should prioritize stratified or population-based sampling to ensure broader applicability of the results.

Conclusions

In summary, this cross-sectional study aimed to investigate the factors influencing breakthrough COVID-19 infections among vaccinated individuals. The findings revealed significant associations between specific community activities and breakthrough infections in vaccinated individuals, underscoring the importance of targeted interventions. Particularly noteworthy was the consistently higher adherence to social distancing and mask-wearing across various settings among the 18 to 29 years age group, providing valuable insights for the formulation of effective public health strategies. However, the study did not identify significant associations between breakthrough infections and demographic variables, vaccine types, or specific activities related to social distancing and mask-wearing. This underscores the nuanced and multifaceted nature of breakthrough infections, suggesting that individual behaviors and contextual factors play crucial roles. Despite the inherent limitations and challenges in predicting breakthrough infections, this study contributes to our understanding of the complex interplay between vaccination status, community behaviors, and infection risks. The findings highlight the ongoing need for research and flexible public health strategies to adapt to the evolving landscape of the COVID-19 pandemic. As we transition toward endemicity, the role of vaccines becomes pivotal in preventing severe outcomes, and the insights from this study can inform future interventions and policies.

Acknowledgments

The authors declare that no generative artificial intelligence tools were used for any sections of this research study or manuscript preparation.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MJM
Data curation: BNV
Formal analysis: BNV
Methodology: MJM, BNV
Project administration: MJM
Validation: MJM, BNV
Visualization: BNV
Writing – original draft: MJM, BNV
Writing – review & editing: MJM, BNV

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia appendices 1
STROBE Checklist
[DOCX File, 37 KB - ojphi_v18i1e76679_app1.docx]

References

1. COVID-19 by county. Centers for Disease Control and Prevention. URL: https://archive.cdc.gov/www_cdc_gov/coronavirus/2019-ncov/your-health/covid-by-county.html? [accessed 2026-03-03]
2. Umakanthan S, Sahu P, Ranade AV, et al. Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). Postgrad Med J 2020 Dec;96(1142):753-758. [doi: 10.1136/postgradmedj-2020-138234] [Medline: 32563999]
3. Durán-Polanco L, Siller M. Crowd management COVID-19. Annu Rev Control 2021;52:465-478. [doi: 10.1016/j.arcontrol.2021.04.006] [Medline: 33867812]
4. Wong R, Grullon JR, Lovier MA. COVID-19 risk factors and predictors for handwashing, masking, and social distancing among a national prospective cohort of US older adults. Public Health 2022 Oct;211:164-170. [doi: 10.1016/j.puhe.2022.08.002] [Medline: 36152400]
5. Rahimi F, Talebi Bezmin Abadi A. Criticality of physical/social distancing, handwashing, respiratory hygiene and face-masking during the COVID-19 pandemic and beyond. Int J Clin Pract 2020 Nov;74(11):e13656. [doi: 10.1111/ijcp.13656] [Medline: 32772412]
6. Howard J, Huang A, Li Z, et al. An evidence review of face masks against COVID-19. Proc Natl Acad Sci USA 2021 Jan 26;118(4):e2014564118. [doi: 10.1073/pnas.2014564118] [Medline: 33431650]
7. Guo Z, Zeng T, Lu Y, et al. Transmission risks of Omicron BA.5 following inactivated COVID-19 vaccines among children and adolescents in China. Commun Med (Lond) 2024 May 18;4(1):92. [doi: 10.1038/s43856-024-00521-y] [Medline: 38762678]
8. Goodman B. First on CNN: CDC set to stop tracking community levels for Covid-19. CNN. 2024.
9. Panagiotakopoulos L, Moulia DL, Godfrey M, et al. Use of COVID-19 vaccines for persons aged ≥6 months: recommendations of the advisory committee on immunization practices - United States, 2024-2025. MMWR Morb Mortal Wkly Rep 2024 Sep 19;73(37):819-824. [doi: 10.15585/mmwr.mm7337e2] [Medline: 39298394]
10. Buet D, Gallais I, Lauret E, et al. Cotargeting signaling pathways driving survival and cell cycle circumvents resistance to Kit inhibitors in leukemia. Blood 2012 May 3;119(18):4228-4241. [doi: 10.1182/blood-2011-07-368316] [Medline: 22438255]
11. Preventing spread of respiratory viruses when you’re sick. Centers for Disease Control and Prevention. 2025 Aug 18. URL: https://www.cdc.gov/respiratory-viruses/prevention/precautions-when-sick.html [accessed 2026-03-03]
12. Siedner MJ, Harling G, Reynolds Z, et al. Social distancing to slow the US COVID-19 epidemic: longitudinal pretest–posttest comparison group study. PLoS Med 2020 Aug 11;17(10):e1003376. [doi: 10.1371/journal.pmed.1003244] [Medline: 32780772]
13. VoPham T, Weaver MD, Adamkiewicz G, Hart JE. Social distancing associations with COVID-19 infection and mortality are modified by crowding and socioeconomic status. Int J Environ Res Public Health 2021 Apr 28;18(9):9. [doi: 10.3390/ijerph18094680] [Medline: 33924821]
14. Wrzus C, Hänel M, Wagner J, Neyer FJ. Social network changes and life events across the life span: a meta-analysis. Psychol Bull 2013 Jan;139(1):53-80. [doi: 10.1037/a0028601] [Medline: 22642230]
15. Zhang N, Chen X, Jia W, et al. Evidence for lack of transmission by close contact and surface touch in a restaurant outbreak of COVID-19. J Infect 2021 Aug;83(2):207-216. [doi: 10.1016/j.jinf.2021.05.030]
16. Rogawski McQuade ET, Becker L, Stroup SE, et al. Risk factors and titers of COVID-19 infection in a longitudinal statewide seroepidemiology cohort. BMC Infect Dis 2023 Oct 11;23(1):676. [doi: 10.1186/s12879-023-08670-6] [Medline: 37821853]
17. Lalani T, Lee TK, Laing ED, et al. SARS-CoV-2 infections and serologic responses among military personnel deployed on the USNS COMFORT to New York City during the COVID-19 pandemic. Open Forum Infect Dis 2021 Feb;8(2):ofaa654. [doi: 10.1093/ofid/ofaa654] [Medline: 33553482]
18. Doung-Ngern P, Suphanchaimat R, Panjangampatthana A, et al. Case-control study of use of personal protective measures and risk for SARS-CoV 2 infection, Thailand. Emerg Infect Dis 2020 Nov;26(11):2607-2616. [doi: 10.3201/eid2611.203003] [Medline: 32931726]
19. Wang Y, Tian H, Zhang L, et al. Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. BMJ Glob Health 2020 May;5(5):e002794. [doi: 10.1136/bmjgh-2020-002794] [Medline: 32467353]
20. Gandhi M, Beyrer C, Goosby E. Masks do more than protect others during COVID-19: reducing the inoculum of SARS-CoV-2 to protect the wearer. J Gen Intern Med 2020 Oct;35(10):3063-3066. [doi: 10.1007/s11606-020-06067-8] [Medline: 32737790]

XSL•FO
RenderX

21. Harris RJ, Hall JA, Zaidi A, Andrews NJ, Dunbar JK, Dabrera G. Effect of vaccination on household transmission of SARS-CoV-2 in England. N Engl J Med 2021 Aug 19;385(8):759-760. [doi: 10.1056/NEJMc2107717] [Medline: 34161702]
22. Milman O, Yelin I, Aharony N, et al. Community-level evidence for SARS-CoV-2 vaccine protection of unvaccinated individuals. Nat Med 2021 Aug;27(8):1367-1369. [doi: 10.1038/s41591-021-01407-5] [Medline: 34113015]
23. Townsend JP, Hassler HB, Wang Z, et al. The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study. Lancet Microbe 2021 Dec;2(12):e666-e675. [doi: 10.1016/S2666-5247(21)00219-6] [Medline: 34632431]
24. Simonsen L, Clarke MJ, Schonberger LB, Arden NH, Cox NJ, Fukuda K. Pandemic versus epidemic influenza mortality: a pattern of changing age distribution. J Infect Dis 1998 Jul;178(1):53-60. [doi: 10.1086/515616] [Medline: 9652423]
25. Lavine JS, Bjornstad ON, Antia R. Immunological characteristics govern the transition of COVID-19 to endemicity. Science 2021 Feb 12;371(6530):741-745. [doi: 10.1126/science.abe6522] [Medline: 33436525]

Abbreviations

CDC: Centers for Disease Control and Prevention
MCPHS: Massachusetts College of Pharmacy and Health Sciences
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by E Mensah; submitted 28.Apr.2025; peer-reviewed by C Obianyo, O Obadimu, R Nair, S Penta; revised version received 12.Jan.2026; accepted 27.Jan.2026; published 01.Apr.2026.

Please cite as:
McDonald MJ, Vajravelu BN
Investigation of Community Behaviors, Socioeconomic Factors, and Breakthrough COVID-19 Infections Among Vaccinated Individuals: Cross-Sectional Study
Online J Public Health Inform 2026;18:e76679
URL: https://ojphi.jmir.org/2026/1/e76679
doi:10.2196/76679

© Matthew J McDonald, Bathri Narayan Vajravelu. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 1.Apr.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Fairness Correction in COVID-19 Predictive Models Using Demographic Optimization: Algorithm Development and Validation Study

Naman Awasthi[1], MS; Saad Abrar[1], MS; Daniel Smolyak[1], PhD; Vanessa Frias-Martinez[2], PhD

[1]Department of Computer Science, University of Maryland, 8125 Paint Branch Ave, College Park, MD, United States

[2]College of Information Studies, University of Maryland, College Park, MD, United States

Corresponding Author:
Naman Awasthi, MS
Department of Computer Science, University of Maryland, 8125 Paint Branch Ave, College Park, MD, United States

Abstract

Background: COVID-19 forecasting models have been used to inform decision-making around resource allocation and intervention decisions, such as hospital beds or stay-at-home orders. State-of-the-art forecasting models often use multimodal data, including mobility or sociodemographic data, to enhance COVID-19 case prediction models. Nevertheless, related work has revealed under-reporting bias in COVID-19 cases as well as sampling bias in mobility data for certain minority racial and ethnic groups, which affects the fairness of COVID-19 predictions across racial and ethnic groups.

Objective: This study aims to introduce a fairness correction method that works for forecasting COVID-19 cases at an aggregate geographic level.

Methods: We use hard and soft error parity analyses on existing fairness frameworks and demonstrate that our proposed method, Demographic Optimization (DemOpts), performs better in both scenarios.

Results: We first demonstrate that state-of-the-art COVID-19 deep learning models produce mean prediction errors that are significantly different across racial and ethnic groups at larger geographic scales. We then propose a novel debiasing method, DemOpts, to increase the fairness of deep learning–based forecasting models trained on potentially biased datasets. Our results show that DemOpts can achieve better error parity than other state-of-the-art debiasing approaches, thus effectively reducing the differences in the mean error distributions across racial and ethnic groups.

Conclusions: We introduce DemOpts, which reduces error parity differences compared with other approaches and generates fairer forecasting models compared with other approaches in the literature.

(Online J Public Health Inform 2026;18:e78235) doi:10.2196/78235

KEYWORDS

COVID-19 forecasting; deep learning model; fairness; time series model; regression

Introduction

Background

Forecasting the number of COVID-19 cases, hospitalizations, or deaths is crucial to inform decision-making. For example, COVID-19 forecasts can be used by hospitals to evaluate medical needs and required resources, such as supplies or beds, or by public health officials to inform closure policies at various geographical scales. In the United States, COVID-19 forecasts have been used at the state and county levels to inform social distancing or masking, such as the publicly available forecasts on the COVID-19 Forecast Hub that the Centers for Disease Control and Prevention (CDC) has routinely used in their communications [1,2].

Related work over the past 4 years has shown a diverse variety of COVID-19 forecasting approaches [3-10, 11] using datasets such as the New York Times (NYT), Johns Hopkins University, COVID-19 Community Vulnerability Index, Google, and Apple [12-16], among others. Most publications focused on COVID-19 case prediction have reported results around the accuracy of the models, that is, minimizing the difference between the predicted cases and the actual number of cases reported. Nevertheless, previous work has shown that the accuracy of COVID-19 predictions can depend on various social determinants, including race or ethnicity [17], income, or age [18], revealing worse performance for protected attributes and pointing to a lack of COVID-19 predictive fairness that can affect resource allocation and decision-making. This lack of predictive fairness might be related to bias in the datasets used to train the model, that is, bias in COVID-19 case reporting [19] or bias in mobility data [20].

XSL•FO
RenderX

Given the presence of bias in the training datasets frequently used by COVID-19 forecast models, and previous work demonstrating that COVID-19 prediction accuracy can vary across social determinants, it becomes critical to devise methods to prevent data biases from percolating into the COVID-19 forecasts to guarantee fair decision-making based on case predictions. In this paper, we focus on in-processing bias mitigation approaches given their scarcity in the COVID-19 literature and propose Demographic Optimization (DemOpts), a debiasing method designed to achieve COVID-19 case prediction error parity across racial and ethnic groups in the context of deep learning models, that is, guarantee that county prediction errors are not significantly different across racial and ethnic groups. Although there exists a diverse set of COVID-19 predictive approaches, we focus on deep learning models because these are the most frequently used models in the machine learning community [21], and narrow down our choice to transformer-based architectures because they are state-of-the-art in time series predictions [22].

The main objective of DemOpts is to improve the fairness of the COVID-19 case predictions at the county level by achieving error parity in a regression setting [17]. DemOpts proposes a novel debiasing approach that leverages county racial and ethnic data during training to modify conventional deep learning loss functions to penalize the model for statistically significant associations between the predictive error and the race or ethnicity distribution of a county. Our main contributions are:

- We present DemOpts, a novel debiasing method for deep learning architectures that attempts to increase the fairness of the COVID-19 county case predictions by achieving error parity, that is, guaranteeing that prediction errors are similar across racial and ethnic groups. The DemOpts architecture is designed to optimize error parity across race and ethnicity using a novel multilabel approach that allows each county to be characterized by its own racial and ethnic group distribution during the debiasing process, instead of by a unique label.
- We propose a novel evaluation protocol for the COVID-19 context, and we show that (1) state-of-the-art COVID-19 county case prediction models based on transformer architectures with no debiasing approach lack error parity, that is, prediction errors are statistically significantly different across racial and ethnic groups, (2) DemOpts applied to transformer-based architectures improves the error parity of the prediction models, increasing the similarity between mean prediction errors across racial and ethnic groups, and (3) the DemOpts debiasing approach performs better than state-of-the-art debiasing methods for regression settings.

While COVID-19 research was particularly prominent from 2020 to early 2024, challenges related to data biases and sampling issues in predictive modeling remain highly relevant. Our approach, leveraging the regression fairness model DemOpts, provides a robust framework to address these challenges. As future pandemics and public health crises arise, similar issues will persist, making our contribution valuable for ensuring fairness and reliability in predictive models.

Literature Review

Deep LearningBased Forecasting Models

Deep learning models have started to become popular in time series prediction tasks. The available methods include (1) autoregressive models, such as Long Short-Term Memory or Gated Recurrent Network [23]; (2) graph-based neural networks, such as graph attention networks [24], Spatio-temporal Graph Convolutional Network [25], neighbor convolution model [26], or graph convolutional network; and (3) transformers, including Logarithmic Sparse Transformer [27], Informer [28], Autoformer [29], Frequency Enhanced Decomposed Transformer [30], Pyramidal Attention–based Transformer [31], and Patch Time Series Transformer [32]. In this paper, we specifically focus on the temporal fusion transformer (TFT) architecture [22], since it allows us to easily incorporate exogenous variables (eg, mobility data) as well as static variables (eg, demographic data) on top of the COVID-19 time series.

Bias in Mobility and COVID-19 Data

The COVID-19 epidemic was closely monitored and had extensive data available about the counts of cases, hospitalizations, and deaths, as well as fine-grained information about mobility of people, policy implementations, vaccinations, and so on. Reducing the impact of mobility data or COVID-19 case bias in COVID-19 case predictions, as we do in this paper, is of critical importance to support decision-making processes focused on resource allocation during pandemics, to reduce harm and guarantee that decisions are fair and just across racial and ethnic groups. Human mobility data has been used to characterize human behaviors in the built environment [33-37], for public safety [38,39], during epidemics and disasters [40-45], as well as to support decision-making for socioeconomic development [46-53]. During the COVID-19 pandemic, human mobility has played a central role in driving decision-making, acknowledging the impact of human movement on virus propagation [7,9,10,18,54]. Previous work has revealed sampling bias in mobility data collected via mobile apps, with Black and older individuals being underrepresented in the datasets [20], and has exposed biases in COVID-19 forecasting models [55,56]. COVID-19 underreporting bias has been discussed in the literature [57-59] and points to multiple causes, including inadequate testing across certain minority groups or a lack of consistency in reporting race and ethnicity for COVID-19 cases [19].

Fairness Metrics and Fairness Corrections

Transformer-based COVID-19 case forecast models require the use of fairness metrics for regression settings, given that the loss optimization process in gradient-based deep learning architectures uses real-number predictions instead of classes.

Agarwal et al [60], Fitzsimons et al [61], and Gursoy and Kakadiaris [17] outline the different aspects of fairness in regression settings and propose a set of fairness metrics for regression-type models. For this paper, we use the error parity metric proposed in [17]. Error parity requires error distributions to be statistically independent of racial and ethnic groups. We expand this definition and relax the statistical significance

XSL•FO
RenderX

requirement to be able to also evaluate whether the proposed DemOpts method can at least reduce the differences in error distributions across racial and ethnic groups, even when they are still statistically significantly different. To correct for bias and unfair performance in deep learning models, researchers have used preprocessing [62,63] and in-processing correction approaches [64-67]. Preprocessing approaches focus on creating a better input for learning deep neural network models by removing bias from the datasets [62,63], and there have been successful efforts focused on debiasing underreporting COVID-19 datasets to estimate actual cases or deaths before they are fed into predictive models [68,69]. On the other hand, in-processing approaches to improve the fairness of deep learning models, like the one we use in this paper, focus on the model and its regularization, usually adding a bias correction term in the loss function [65,67]. In this paper, we will compare our proposed debiasing approach against 3 state-of-the-art methods for debiasing in regression settings, which are individual fairness correction [70], group fairness correction [70] (both Lagrangian-based), and sufficiency [71]. Individual and group fairness calculate penalties by determining overestimations across different groups and weighting the loss by a factor proportional to the overestimations, while sufficiency-based regularizers propose to make the loss independent of sensitive data attributes by simultaneously training a joint model and subgroup-specific networks to achieve fair predictions [71].

Methods

Proposed DemOpts

Our modeling focus is on deep learning models, which are the most frequently used approach for COVID-19 county case forecasts in the machine learning community [21]. We specifically focus on the TFT model introduced in [22] for several reasons. First, this model is state-of-the-art in interpretable time series prediction [22]. Second, this model allows for the use of static reals as input to the model (ie, attributes that do not change over the duration of the training process, such as demographic percentages or population statistics). Third, the model works well with time-dependent features, including COVID-19 cases or mobility data, whereby past data influences future statistics.

DemOpts is an in-processing algorithm that modifies the standard training procedure for deep learning models at the loss computation stage. The algorithm modifies conventional loss functions to penalize the model for any statistically significant association ($P<.005$) between the county prediction loss (error) and the county's racial and ethnic groups. In other words, DemOpts performs a race-based correction on the error to account for county demographic, racial, and ethnic distributions.

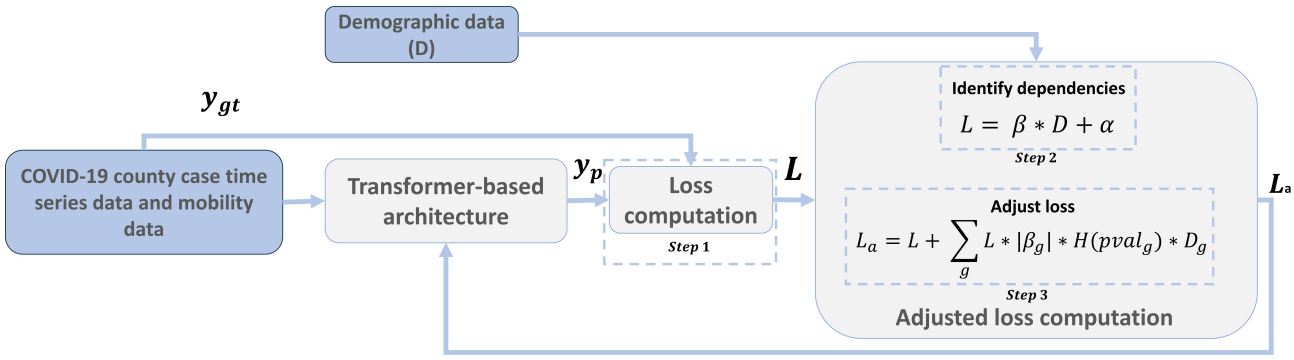
The algorithm can be divided into 3 steps (refer to Figure 1, Figure 2, and "S.1 DemOpts Method" in Multimedia Appendix 1 for mathematical details).

XSL•FO
RenderX

Figure 1. Algorithm: Demographic Optimization (DemOpts). TFT: Temporal Fusion Transformer.

```
Input: Training set (X, D, Y), Learning rate (lr), Number of epochs, threshold
Output: Trained model (M)
X: COVID-19 Timeseries data for all counties
Y: COVID-19 cases in future for all counties
D: Demographic data for all counties
Initialize model parameters randomly
for epoch in range(0, epochs) do
    // sample from X, D, Y of size b
    for (X_b, D_b, Y_bt) in (X, D, Y) do
        // Forward propagation
        Y_bp = M(X_b)
        //Calculate QuantileLoss
        L_i = QuantileLoss(Y_bp, Y_bt)
        //Find association
        olsreg = OLS.fit(D_b, L_b)
        pvals, β = olsreg.pvals, olsreg.coef
        // additional penalty on loss
        for index in |pvals| do
            pval_i, β_i = pvals[index], β[index]
            // Get the corresponding demographic percentage column and all rows
            D_b,idx = D_b[:, index]
            if pval_i < threshold then          //this ensures significant association
                L_b += L_b * |β_i| * D_b,idx
            end if
        end for
        backpropagate(M, L_b)
    end for
end for
return TFT
```

Figure 2. Flow diagram for the Demographic Optimization (DemOpts) method.



Step 1: Calculate Loss

We use quantile predictions, as standard in COVID-19 forecasting literature [2,72], instead of point-value predictions. Quantile predictions are measured for 7 quantiles (0.02, 0.1, 0.25, 0.5, 0.75, 0.9, and 0.98) to gain insights into the uncertainty ranges and CIs of the COVID-19 county case predictive models. When using quantile predictions, the error is computed using quantile loss, also known as pinball loss (PBL), and defined as follows:

PBLq(yip,yi)={q*(yi−yip) if yi≥yip(q−1)*(yi−yip) if yi<yip

For quantile q, the PBL for the prediction of a given input X_i is $PBL_q(y_{ip}, y_i)$, where y_i is the ground truth, and y_{ip} is the predicted value. The average over all quantiles can be represented as PBLyip, yi = 1q∑qPBLqyip, yi.

Step 2: Identify Dependencies Between Prediction Errors, Race, and Ethnicity

To achieve error parity, that is, mean errors being independent of racial and ethnic population distributions, we determine the relationship between errors and race and ethnic distributions. For that purpose, DemOpts fits a regression model between the

XSL•FO
RenderX

prediction losses PBL (y_{ip},y_i) across data points *i* and their corresponding county race and ethnicity distribution for each race D_i:

$$PBL(yip,yi)= \beta*Di + \alpha \text{ with } Di = [d1,d2,d3,d4, lookahead]$$

where d_i are the corresponding county demographic features extracted from the US census data (represented as the percentage of each racial and ethnic group of the county for datapoint i), and lookahead refers to the number of days into the future the COVID-19 case prediction was generated for. In matrix representation:

$$PBLYip,Yi=\beta*D+\alpha$$

Once the regression model is fit, both regression coefficients (β) and their statistical significance (*P* value) are passed on to Step 3 to modify the adjusted loss and attempt to decouple race from the errors (loss).

Step 3: Adjust the Loss

DemOpts modifies the conventional loss of deep learning models by adjusting for racial or ethnic bias in the error, that is, the loss is increased whenever a statistically significant regression coefficient for a race or ethnicity is found in Step 2 (with *P* value threshold=.005). By increasing the loss, DemOpts attempts to reduce the association between errors and race. Specifically, the loss is adjusted by the product of the original loss PBL (y_{ip},y_i), the percentage race or ethnicity D_j that holds a significant relationship with the error, and its coefficient β_j in absolute value:

$$Ladj=PBL(yip,yi)+\sum jH(pvalj)(|\beta j|*Dj*L) \text{ where}$$

$$H(x)=\{1 \quad \text{if } x<0.0050 \quad \text{if } x\geq 0.005$$

Evaluation Protocol

In this section, we present a novel evaluation protocol to assess changes in fairness for TFT forecasting models when debiasing approaches, including DemOpts, are applied. We first describe the TFT COVID-19 county case prediction model we use, and the different debiasing approaches we evaluate on that prediction model. Next, we describe the error parity metrics we use to evaluate the fairness of each prediction model, and finally, we present the approach to analyze whether DemOpts improves the error parity metrics when compared to other state-of-the-art debiasing approaches for regression settings.

Predictive Model and Debiasing Approaches

We use the TFT with the conventional PBL function (PBL is the standard metric for reporting model performance in CDC Forecast Hub [2]) as our baseline model ($TFT_{Baseline}$) to predict the number of COVID-19 county cases for a given day.

Input data to the TFT model includes past COVID-19 cases per county, mobility data from SafeGraph, and race and ethnicity data for the county. We also train and test another TFT enhanced with the DemOpts debiasing method, $TFT_{DemOpts}$, that adjusts the loss computation to attempt to eliminate or reduce the dependencies between error and race to achieve error parity. In addition, we train and test 3 more TFTs enhanced with state-of-the-art debiasing methods for regression settings, namely, individual fairness $TFT_{Individual}$ [70], group fairness TFT_{Group} [70], and the sufficiency-based regularizer $TFT_{Sufficiency}$ [71]. Individual and group fairness methods calculate penalties by determining overestimations across different groups and weighting the loss by a factor proportional to the overestimations, while the sufficiency-based regularizer trains a joint model and group-specific networks to achieve fair predictions. We replicate their methodology and adapt it to the forecasting setting by keeping TFT as the common network.

Measuring Model Fairness

We choose error parity as our fairness metric [17], with a focus on evaluating whether the distribution of predictive errors at the county level is independent of county majority race and ethnicity, that is, prediction errors are not statistically significantly different across racial and ethnic groups. To measure the fairness of each of the models $TFT_{Baseline}$, $TFT_{DemOpts}$, $TFT_{Individual}$, $TFT_{Group}$and $TFT_{Sufficiency}$, we propose a 2-step process.

Step 1: Associate Errors With County Race or Ethnicity

To carry out the fairness analysis, we need to associate the PBL error of each county with race and ethnicity labels. However, that would require access to race-stratified COVID-19 case data at the county level, which is unfortunately not available due to systemic data collection failures during the pandemic [73]. Hence, we propose to associate each county and its error with the majority race, that is, we label each county with the race or ethnicity that has the highest population percentage in that county. During the fairness analysis, we refer to majority White counties as the unprotected group and majority minority counties, such as Black or Hispanic, as the protected groups (details about the racial and ethnic groups considered in the evaluation are provided in the “Datasets” section).

In addition, we normalize each county’s PBL error by county population size. The normalization by county population allows us to scale the errors appropriately, since higher-population counties will have higher case counts and thus, higher-magnitude errors. Normalizing by population fairly compares the error per unit population of one county with another:

$$NormPBLypi,yti=1000\cdot PBLyp,i,yt,ipopi$$

where y_{ti} is the ground truth, y_{pi} is the predicted value, and pop_i is the county population.

We then calculate the average normalized PBL for each racial or ethnic group:

$$AvgNormPBL(yp,yt,g)=\sum i\in cgNormPBL(ypi,yti)|cg|$$

where g represents the racial or ethnic group and c_g is the set of all counties with as the majority group. This gives us the average normalized PBL for each demographic group.

Step 2: Compute Fairness Metric

Once PBLs have been calculated for each racial and ethnic group in the United States, we can compute the error parity, that is, the fairness metric focused on evaluating whether the prediction errors are different across race and ethnicity. We

XSL•FO
RenderX

propose 2 metrics to measure the error parity of COVID-19 county case predictions: hard error parity and soft error parity.

Hard Error Parity Metric

Model predictions exhibit hard error parity when no statistically significant differences exist between normalized mean case prediction errors (AvgNormPBL) across racial or ethnic groups. In other words, normalized mean PBL errors across counties of different racial and ethnic groups are similar and hence, not biased by race or ethnicity. To test for the hard error parity of a prediction model, we propose to run one-way ANOVA followed by post hoc Tukey honestly significant difference (HSD) tests between the normalized mean error distributions of all racial and ethnic groups. ANOVA tests are an adequate choice even in violation of normality for large sample sizes, and in the presence of unequal sample sizes with homogeneous variance; thus, we choose this parametric test due to its superior strength [74,75].

Rejecting the null hypothesis for ANOVA would point to significantly different mean error values across some racial or ethnic groups and to a lack of perfect hard error parity. The subsequent analysis of the post hoc Tukey HSD test would reveal the pairs of racial and ethnic groups whose mean error values are significantly different and the numerical difference. The Tukey test also highlights the pairs of racial and ethnic groups for which the mean error is not statistically significantly different, pointing to instances where hard error parity exists for that model.

Soft Error Parity Metric

Instead of measuring the statistical significance of the relationship between county race labels and county errors, we propose to use the Accuracy Equity Ratio (AER) metric [76]. AER computes the ratio between the errors of the protected and unprotected groups as follows:

$$AERpg=AvgNormPBLyp,yt,pgAvgNormPBLyp,yt,unpg$$

where subscript pg indicates counties labeled as the protected group (majority minority counties). unpg indicates counties labeled as the unprotected group (White), and AvgNormPBL is the average of the normalized PBL across counties for a given racial group g (pg or unpg).

As defined, the AER metric goes from 0 to ∞. AER values in the range [0, 1] indicate comparatively lower normalized PBL for protected groups, which means the model predictions could be biased—have higher errors—for White majority counties; while AER values larger than one indicate that the model could be biased against the protected group, that is, the prediction errors are larger for counties with majority-minority groups. Values close to 1 indicate parity in error distribution between the protected group counties and the majority White counties. We claim that a predictive model achieves soft error parity for a given protected group when the AER value is close to 1, that is, the mean predictive error between that protected group and the White race is similar.

An alternative approach to assigning majority race or ethnicity would be to explore the associations between PBL errors and the distribution of racial and ethnic groups in a county (independent of COVID-19 cases, since that data are not available). Using a quantile regression, we can explore whether DemOpts eliminates significant associations between racial or ethnic percentages and the PBL errors, or at least reduces their magnitude. This approach removes the majority race requirement, but does not allow us to perform analyses with well-established fairness metrics in the literature, such as AER. Results are provided in the Multimedia Appendix 1.

DemOpts Over State-of-the-Art

To assess whether DemOpts is a better debiasing approach than state-of-the-art methods, we need to compare the error parity metrics of the COVID-19 county case prediction model enhanced with the DemOpts method, $TFT_{DemOpts,}$ against the error parity metrics of the same prediction model enhanced with the other debiasing approaches (individual $TFT_{Individual}$, group TFT_{Group}, or sufficiency $TFT_{Isufficiency}$), as well as with the baseline COVID-19 county case prediction model without any debiasing approach, $TFT_{Baseline}$. Next, we describe how we carry out this analysis for the hard and soft error parity metrics.

Hard Error Parity

We computed the hard error parity metric for each of the COVID-19 county case prediction models, using one-way ANOVA and the post hoc Tukey HSD test. An exploration of the statistical significance of the mean error difference for each pair of racial and ethnic groups will reveal whether applying DemOpts to the COVID-19 case prediction model produces fewer instances of significant mean error differences than any of the other debiasing methods or the baseline. In other words, a decrease in the number of significantly different mean PBL errors between races would point to an achievement of hard error parity for more racial and ethnic groups than other state-of-the-art debiasing approaches or the baseline.

Soft Error Parity

To assess whether DemOpts applied to a COVID-19 case prediction model has higher soft error parity than any of the other state-of-the-art debiasing approaches, we propose to compare the AER values for each protected race and ethnic group across the 5 models: $TFT_{DemOpts}$, $TFT_{Individual}$, TFT_{Group}, $TFT_{Sufficiency}$, and $TFT_{Baseline}$. Since AER values represent the quotient between the normalized mean prediction errors of a protected race or ethnicity vs White counties, the model with AER values closer to 1 will be the approach with the highest soft error parity. To measure AER's distance to 1, we compute the distance=$|1\text{-}AER_{race}|$ for each race and ethnic group, which represents the distance to a perfect soft parity error of 1. Distances closer to zero reveal better soft error parities.

Datasets

In this section, we discuss the datasets we use in the DemOpts evaluation in the "Results" section. We train COVID-19 county case prediction models for the United States using COVID-19 case data, as well as mobility and demographic data. Mobility data has been used by previous work to inform case predictions via human mobility behaviors, under the assumption that the way people move might have an impact on the spreading of the epidemic. On the other hand, demographic data, either raw from

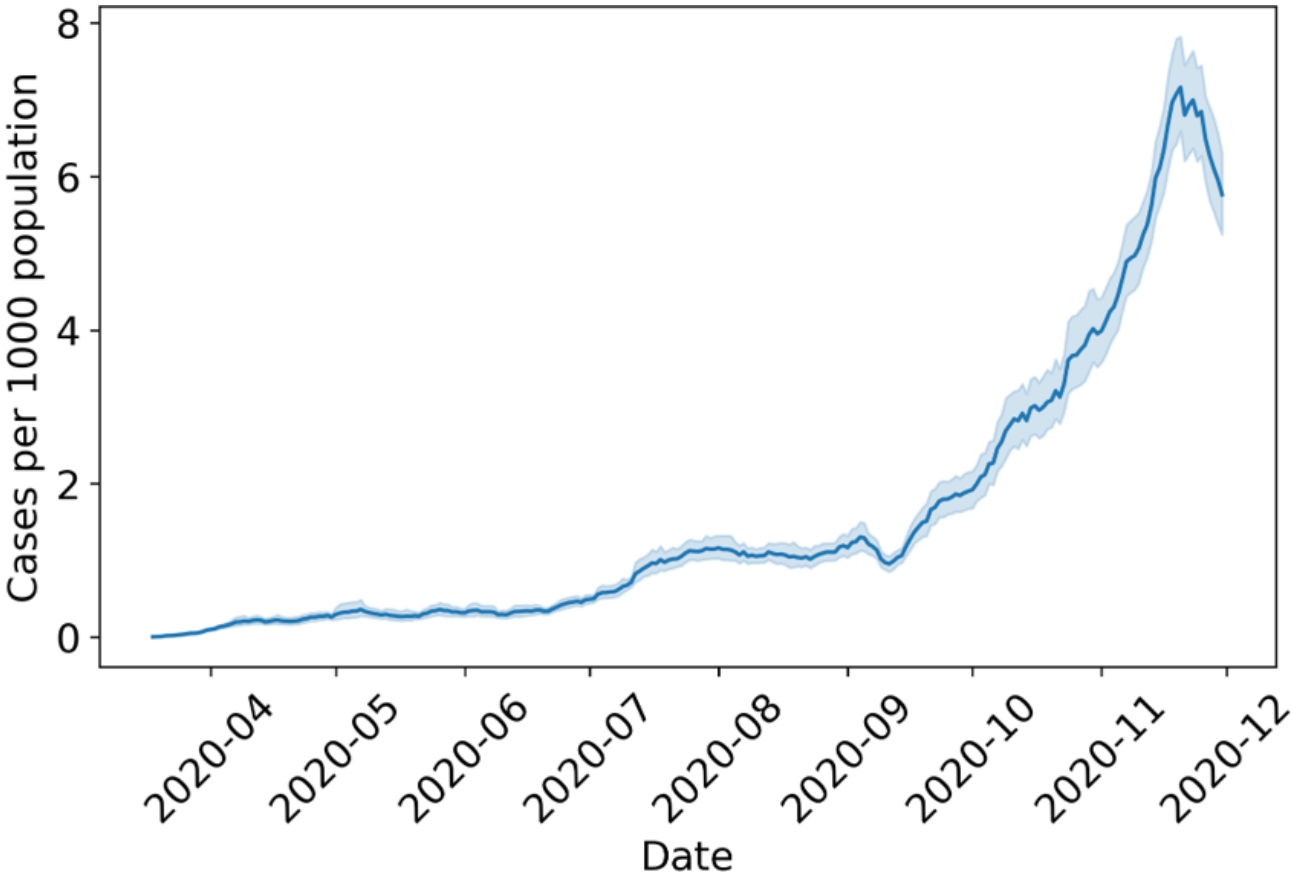
XSL•FO
RenderX

the census or combined in different types of vulnerability indices, has also been shown to help predict COVID-19 prevalence, given the fact that COVID-19 has heavily affected vulnerable populations [59].

COVID-19 Case Data

We use the COVID-19 case data compiled by the NYT at the county level [12]. We account for delayed reporting by using the 7-day daily rolling average of COVID-19 cases (computed as the average of its current value and 6 previous days) instead of raw counts. Figure 3 charts the daily COVID-19 reported cases throughout the data collection period.

Figure 3. COVID-19 reported case counts per 1000 population across the United States.



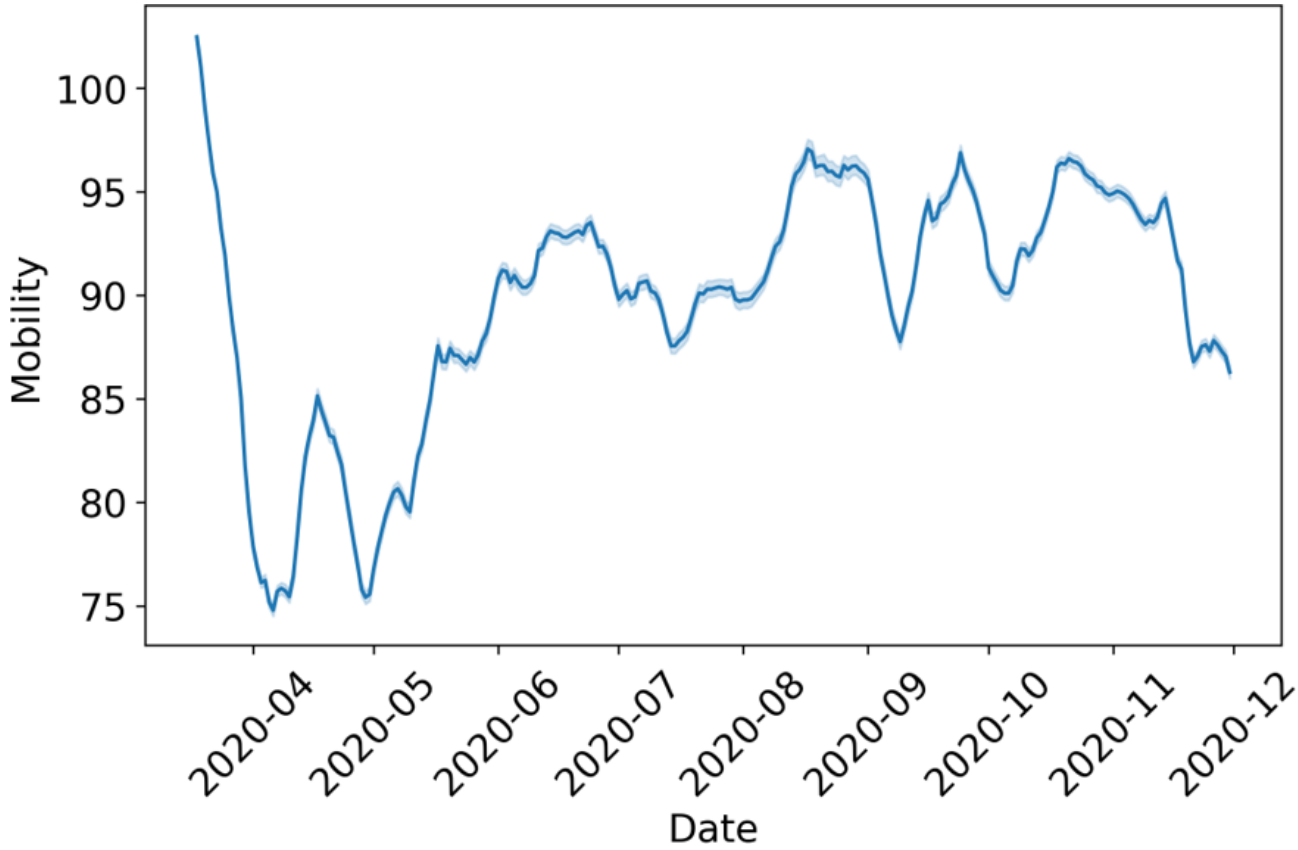
Mobility Data

SafeGraph open-sourced the mobility patterns of smartphone app users at the onset of the pandemic. These data points are curated by tracking the movements of millions of pseudonymized users via mobile app Software Development Kits (SafeGraph). Based on the data available, we use the daily origin-destination (OD) county-to-county flows [77]. OD flows represent the volume of trips between pairs of counties across the United States for each day. For OD flows, we only use SafeGraph inflow (ie, mobility into the county). The inflow mobility is measured as changes in volumes of flows with respect to a baseline of normal behavior computed by SafeGraph using mobility data from February 17, 2020, to March 7, 2020.

Previous work has shown sampling bias in mobility datasets, revealing that not all races and ethnicities are equally represented due to variations in smartphone penetration rates [20,78]. It has also been shown that sampling bias in mobility data can negatively impact downstream tasks such as COVID-19 forecasting [56]. While the addition of mobility data could potentially help improve prediction accuracy and support better decision-making, it also introduces bias. Our empirical analysis of DemOpts aims to understand whether the debiasing method proposed in this paper can improve the fairness of COVID-19 county case predictive models when mobility data is used as input to the predictive model. Figure 4 shows the aggregate mobility data across the country. We see an initial drop in mobility in April (2020‑04), which corresponds to the first lockdown period. We then observed an increase in mobility a month later, which partially stabilizes after April.

XSL•FO
RenderX

Figure 4. Mobility for all ethnic and racial groups.



Race and Ethnicity Data

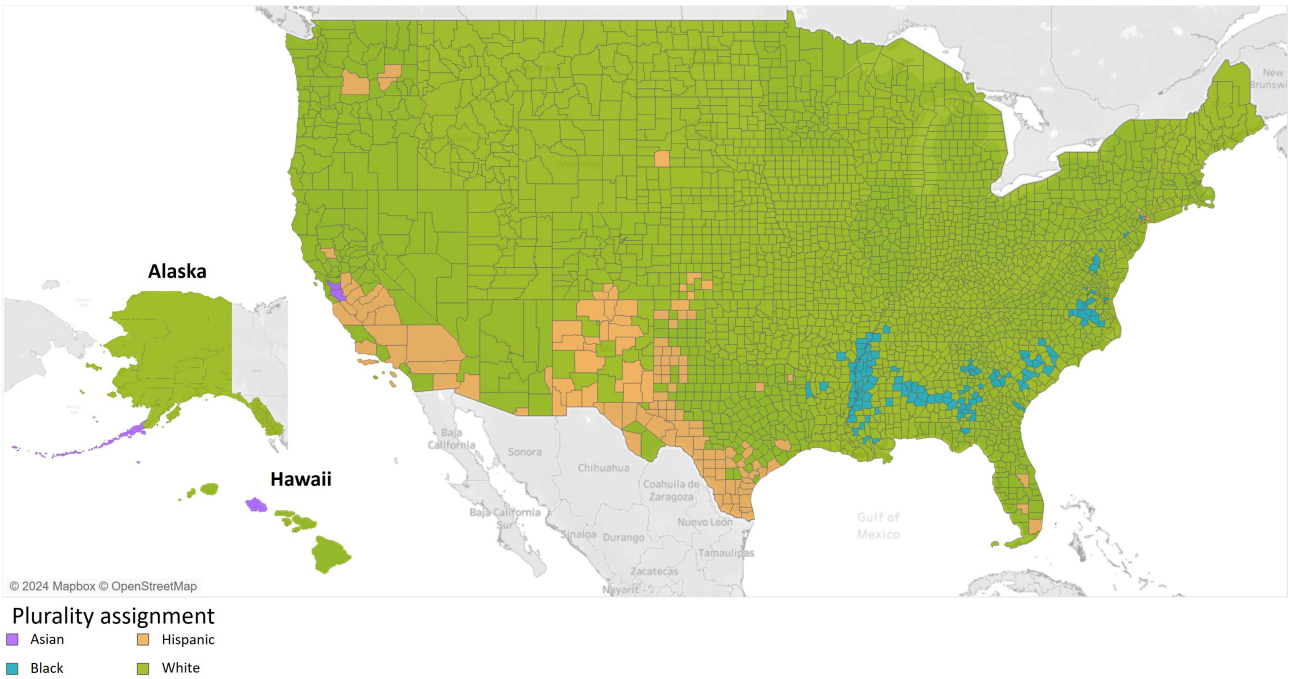
We retrieve the race and ethnicity data from each county in the United States from the 2019 5-year American Community Survey. This survey collects data annually from all 50 states, Puerto Rico, and Washington, DC. As described in Step 1 of the evaluation protocol, we associate each county and its errors with the majority race (ie, we label each county with the race or ethnicity that has the highest population percentage in that county). Following this procedure identifies 4 racial and ethnic groups for the majority of counties: Asian, Black, Hispanic, and White. Table 1 shows the distribution of US counties into these 4 racial and ethnic groups, and Figure 5 show color-coded maps with the majority racial or ethnic group for each county.

Table . Majority label counts.

Majority label	Count, n (%)
Asian	6 (0.194)
Black	127 (4.118)
Hispanic	126 (4.085)
White	2825 (91.601)

XSL•FO
RenderX

Figure 5. Counties and majority-based label. (Mapbox and OpenStreetMap were used to create this map [79,80,81].)



Model Training

For evaluation purposes, we use COVID-19 case and SafeGraph mobility data from March 18, 2020, to November 30, 2020, for the training (207 days) and testing (49 days) of the TFT COVID-19 county case prediction models. The forecast task is the prediction of the number of COVID-19 cases for a given county from day X+1 to X+50, that is, the following 2 months (long-term forecasting with lookahead values from 1 to 50). Specifically, we train and test (1) the $TFT_{Baseline}$, a TFT prediction model without a debiasing method; (2) the $TFT_{Individual}$, TFT_{Group}, and $TFT_{Sufficiency}$, TFT prediction models with state-of-the-art debiasing methods; and (3) $TFT_{DemOpts}$, a TFT prediction model enhanced with our proposed debiasing method. All 5 models are trained and tested for the same temporal range, and all are implemented using the PyTorch Forecasting library [82]. We limit the period of analysis to a time before COVID-19 vaccines were available, given that after that event, research has revealed a less clear relationship between mobility data and postvaccines COVID-19 case volumes [83]. We use the prediction errors (PBL) per racial and ethnic group to analyze and compare the hard and soft error parity of all trained models.

Ethical Considerations

We used openly available datasets for mobility data (SafeGraph), COVID-19 case count (NYT), and demographic data (American Communities Survey). There was no human participant recruitment in this study, and thus we did not require institutional review board approval. All the datasets were aggregated at the county level and do not pose the risk of deanonymization.

Results

Hard Error Parity Results

ANOVA tests of the normalized mean PBL error distributions across racial and ethnic groups for each debiasing approach were all significant, pointing to a dependency between race and the normalized prediction errors.

Table 2 shows the *F* statistic and test significance for each of the prediction models with and without debiasing approaches. The significant ANOVA tests reveal that perfect hard error parity is not achieved by any of the debiasing methods. In other words, for some racial and ethnic groups, there exist statistically significant differences between their mean PBL prediction errors of different racial and ethnic groups; this effect occurs for the $TFT_{Baseline}$ model as well as across all the other predictive models enhanced with a debiasing approach.

Nevertheless, post hoc Tukey HSD tests revealed interesting, nuanced results, showing significant differences in errors only between specific pairs of racial and ethnic groups. Table 3 shows the post hoc Tukey HSD test results for each COVID-19 case predictive model: the baseline, the baseline enhanced with 1 of the 3 state-of-the-art debiasing approaches, and the baseline enhanced with our proposed method (DemOpts). Each row represents the output of the post hoc test, that is, the difference between the normalized mean PBL error of Group 1 and Group 2 ($NormPBL_{Group1}$ - $NormPBL_{Group2}$). If the difference is positive, it means that the normalized mean predictive error is higher for Group 1; if the difference is negative, the normalized PBL error is higher for Group 2 (superscript b indicates statistically significant differences).

The first relevant observation when examining the table is that the baseline model, focused on predicting COVID-19 county cases with no debiasing approach is highly biased, with statistically significant differences between the mean normalized

XSL•FO
RenderX

errors across all pairs of races, except for the comparison between Asian and Black counties as well as Hispanic and White counties, for which there is no statistically significant difference. These results reveal that there is no racial or ethnic group that achieves hard error parity and motivate our exploration of whether state-of-the-art debiasing methods or our proposed DemOpts can improve the hard error parity results of the baseline model.

Table . ANOVA *F* test statistics comparing mean prediction errors.

Fairness method	F statistic (*df*)
Baseline	1195.398[a] (3080)
Group	1455.528[a] (3080)
Individual	1469.698[a] (3080)
Sufficiency	1195.651[a] (3080)
DemOpts[b]	668.769[a] (3080)

[a]*P*<.001.

[b]DemOpts: Demographic Optimization.

Table . Hard error parity analysis. Each value represents the difference between the mean normalized pinball loss for each pair of racial and ethnic groups and indicates whether the difference is statistically significant.

Groups 1 and 2	Baseline	Group	Individual	Sufficiency	DemOpts[a]
Asian					
Black	−0.11	−0.20	−0.12	−0.11	1.32
Hispanic	−2.30[b]	−2.65[b]	−2.50[b]	−2.29[b]	−0.77[c]
White	−2.06[b]	−2.51[b]	−2.51[b]	−2.06[b]	−0.96[c]
Black					
Hispanic	−2.18[b]	−2.45[b]	−2.38[b]	−2.17[b]	−2.09[b]
White	−1.94[b]	−2.31[b]	−2.39[b]	−1.94[b]	−2.29[b]
Hispanic					
White	0.23	0.14	−0.01	0.23	−0.19

[a]DemOpts: Demographic Optimization.

[b]*P*<.001.

[c]These values denote no significant difference between the prediction errors of Asian and White counties and of Asian and Hispanic counties.

When examining Table 3, we can observe that predictive models enhanced with the individual, group, or sufficiency debiasing methods do not improve the hard error parity over the baseline. On the one hand, similarly to the baseline model, the state-of-the-art debiasing methods ($TFT_{Individual}$, TFT_{Group}, and $TFT_{Sufficiency}$) achieve hard error parity between Asian and Black counties and between Hispanic and White counties, that is, the mean error difference between these counties is not significant, pointing to a fair distribution of errors. On the other hand, for each pair of racial and ethnic groups whose prediction error distributions are significantly different for the baseline (rows with asterisks in the Baseline column), they remain significantly different for the individual, group, and sufficiency debiasing methods (rows with superscript b in the individual, group, and sufficiency columns).

When examining the significant mean PBL differences between racial and ethnic groups for the baseline and the state of the art debiasing models, we observe that all coefficients have similar values, signaling similar significant mean PBL differences between racial and ethnic groups (with values between 1.942 and 2.659 error cases per 1000 population). The sign of the coefficients reveals higher mean PBL errors for Hispanic and White counties when compared to Asian or Black counties, and higher mean PBL errors for White counties when compared to Hispanic counties across all models. For example, Hispanic and White counties have mean prediction errors 2.302 and 2.064 cases higher, respectively, when compared to Asian counties and while using the baseline model; and Hispanic and White counties have errors 2.457 and 2.313 cases higher, respectively, when compared to Black counties using the baseline model enhanced with the Group debiasing approach.

Moving on to DemOpts, the table shows that our proposed approach is the only debiasing method that achieves hard error parity in more cases than the baseline, effectively removing some of the associations between race and ethnicity and the normalized mean error distribution (PBL). Specifically, DemOpts removes the significant difference between the prediction errors of Asian and White counties and of Asian and

XSL•FO
RenderX

Hispanic counties (refer to values with supercript c in Table 3), effectively achieving hard error parity for Asian counties, that is, the mean PBL in Asian counties is always similar to the mean error in counties of all the other racial and ethnic groups. These improvements occur additionally to hard error parity already seen in $TFT_{Baseline}$ (hard error parity between Asian and Black counties and between Hispanic and White counties), which are also present in the other 3 debiasing methods. In other words, DemOpts improves the hard error parity of case predictions for 2 additional racial and ethnic pairs compared with any of the other debiasing methods.

Finally, when looking specifically at the hard error parity between protected (Asian, Black, and Hispanic) and unprotected groups (White), DemOpts achieves hard error parity for Asian and Hispanic groups; that is, their mean prediction errors are not significantly different from those of White counties, while the baseline and the other 3 debiasing methods only achieve hard error parity for the Hispanic group when compared to White counties. These findings with respect to White counties motivate the evaluation of the soft error parity of the different models to determine, for example, whether DemOpts achieves the best soft error parity for the Black group (since hard error parity was not achieved), or to see if DemOpts has better soft error parity than other debiasing methods for Asian or Hispanic groups. Next, we explore the soft error parity metric for the TFT baseline and for all TFT models enhanced with debiasing approaches.

Soft Error Parity Results

Table 4 shows the distance to the perfect soft error parity for each of the debiasing approaches across all protected racial and ethnic groups. As we can observe, DemOpts has the smallest values—closest distances to perfect soft error parity—for Asian and Black counties, while the individual debiasing method almost achieves perfect soft error parity for the Hispanic counties. In other words, DemOpts is the debiasing approach that produces the most similar errors between Asian and White counties and between Black and White counties, thereby achieving the largest reduction in predictive bias. On the other hand, the Individual debiasing method achieves errors for Hispanic counties that are closest to the White group. In addition, it is important to highlight that the Group and Sufficiency debiasing methods achieve soft error parities that are close to the $TFT_{Baseline}$, which is not enhanced with any debiasing method.

Table . Soft error parity analysis. Each value represents the distance ($|1\text{-}AER_{race}|$) for each protected group and debiasing method. $TFT_{DemOpts}$ achieves the highest soft error parity for 2 of the 3 protected races under study.

Group	Baseline	Group	Individual	Sufficiency	DemOpts[a]
Asian	0.811	0.842	0.850	0.811	0.454[b]
Black	0.764	0.774	0.807	0.764	0.681[b]
Hispanic	0.093	0.048	0.003[b]	0.093	0.12

[a]DemOpts: Demographic Optimization.

[b]Smallest error parity for the particular group

Overall, these results reveal that DemOpts is the debiasing approach that improves the soft error parity of case prediction models, with errors for Asian and Black counties being the closest to errors in White counties. When accounting for additional factors, DemOpts outperforms the other methods by reducing the racial associations of model error.

In Table S1 in Multimedia Appendix 1, we provide and discuss the results for the quantile regression analysis in detail. Overall, the results confirm our findings with majority race labels, with DemOpts consistently outperforming other methods, showing the smallest coefficient magnitude for associations between the percentage of Asian, Black, and Hispanic populations and model error.

Discussion

Principal Findings

Through our comparison of model performance for COVID-19 case prediction across counties of differing racial demographics, we showed that DemOpts outperforms other baselines for debiasing predictions. In our analysis of hard error parity, we found that DemOpts was the only debiasing method to eliminate statistically significant relationships between prediction error and racial demographics when compared with the baseline. While some significant associations remained, DemOpts achieved hard error parity for Asian vs White counties and Asian vs Hispanic counties. In the soft error parity analysis, DemOps substantially outperformed the baselines for Asian and Black counties, with a 69.4% reduction and 23% reduction, respectively, compared with the next closest method.

Why is DemOpts Better?

The results showed that DemOpts is the only debiasing approach to achieve both hard and soft error parity for all 3 racial minority groups when compared with White counties.

In an attempt to understand why DemOpts succeeds in increasing both hard and soft error parity in the context of COVID-19 county case predictions, and compared with other debiasing methods, we computed the average PBL for each racial and ethnic group and for each predictive model enhanced, or not, with a debiasing method (refer to Table 5). We observed that DemOpts achieves better hard and soft error parity metrics because it considerably increases the errors for Asian and Black counties with respect to the baseline, until the differences with Hispanic and White are made not statistically significant (hard error parity) or closer to the White mean errors (soft error parity). Comparing Tables 4 and 5, we observed that DemOpts

XSL•FO
RenderX

achieves considerably higher fairness for the Hispanic group (when compared to White) than for the Asian and Black groups (0.12 vs 0.454 and 0.681 in Table 4). As a result, the average PBL error for the Hispanic group (3.59 in Table 5) is considerably higher than the Asian and Black racial groups (1.7 and 1, respectively). We hypothesize that the differences in average errors and performance across racial and ethnic groups could be due to differences in the bias present in the training data, that is, mobility data or COVID-19 case counts could be more biased for Asian or Black groups, thus making it harder to achieve fair predictions when compared to White, and, in turn, due to the fairness-accuracy trade-off, making them more accurate (lower errors).

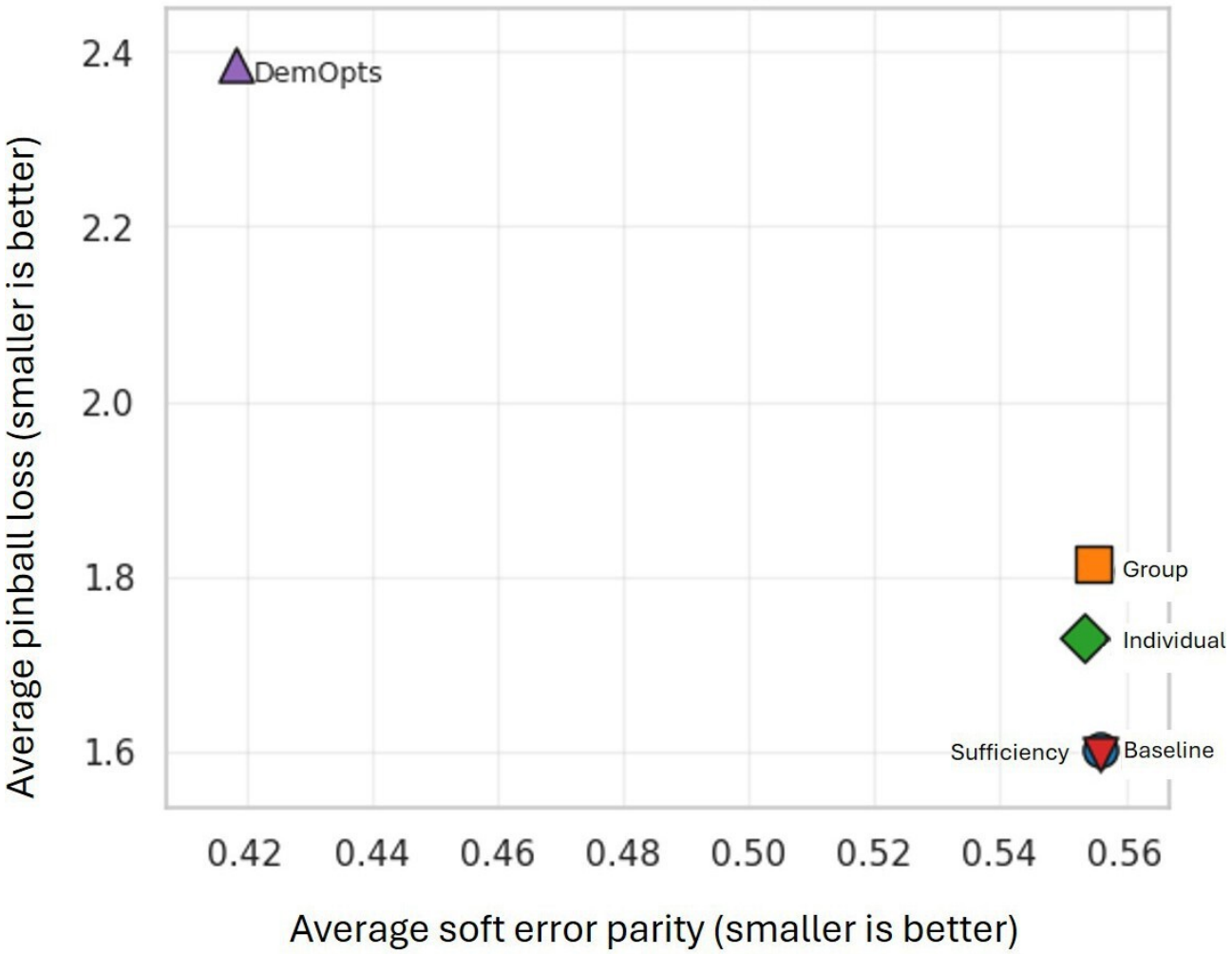
Table . Group-wise pinball loss for each model. Demographic Optimization (DemOpts) has higher average pinball loss compared to the other models. The fairness-accuracy tradeoff leads to slightly larger pinball loss values for DemOpts compared to other methods.

Group	Baseline	Group	Individual	Sufficiency	DemOpts[a]
Asian	0.482	0.472	0.444	0.479	1.741
Black	0.600	0.674	0.570	0.598	1.015
Hispanic	2.784	3.131	2.951	2.776	3.597
White	2.546	2.987	2.961	2.540	3.192

[a]DemOpts: Demographic Optimization.

These results show that DemOpts' optimization could not decrease prediction errors while trying to improve fairness, showing a fairness-accuracy trade-off that has been reported previously in the literature [84]. To further clarify this finding, Figure 6 shows both the average PBL and soft parity across all the models considered in this paper. As shown, DemOpts has the lowest soft error parity, but the highest PBL (top-left corner in the plot), while the other models decrease the PBL by sacrificing fairness (higher error parity in the bottom-right corner).

Figure 6. Fairness-accuracy tradeoff. Model error (average pinball loss) vs average soft error parity (|1-AER|) for each model.



XSL•FO
RenderX

Limitations

While DemOpts outperforms other state-of-the-art approaches in debiasing COVID-19 predictions, there are some limitations to DemOpts and our evaluation. First, DemOpts is unable to remove all statistical associations for the hard parity analysis, potentially because doing so would impose further reductions in model performance. For the soft parity analysis, the individual fairness approach is best for Hispanic counties, but the difference in parity levels is small. Regarding evaluation, our focus is exclusively on COVID-19 county case prediction—while evaluation on other datasets and prediction tasks would be helpful for future work, our current evaluation provides sufficient evidence to show its applicability to other contexts. In addition, we compare DemOpts to baselines only on error parity metrics. Other fairness metrics may apply to the COVID-19 context and should be evaluated in future work, but we focus on error parity because DemOpts is specifically designed to mitigate it. Finally, we only compare DemOpts and baseline debiasing approaches within TFT models—future work should compare with other commonly used models for COVID-19 case prediction.

Regardless, our novel debiasing approach shows that hard and soft error parity across protected and unprotected racial and ethnic groups can improve relative to other state-of-the-art approaches.

Finally, it is important to clarify that although in this paper, DemOpts focuses on bias mitigation in COVID-19 forecasting, it could also be applied to other health forecasting tasks where sampling bias in data collection can lead to bias in downstream tasks, for example, forecasting flu cases. These forecasts, when done at the county level and when using mobility data to model human spread, could benefit from the DemOpts method by reducing the effect of mobility bias or case count bias on other infectious diseases.

Conclusion

Researchers have worked tirelessly on the creation of accurate COVID-19 case prediction models to support resource allocation and decision-making. However, sampling and underreporting biases in the data used to train these models have resulted in worse prediction performance for certain protected attributes, pointing to a lack of COVID-19 predictive fairness that could affect decision-making. In this paper, we show that state-of-the-art architectures in COVID-19 case predictions (TFT models) incur unfair prediction error distributions, and we design a novel debiasing approach and evaluation method to increase the fairness of predictions in the context of COVID-19 county case forecasts. DemOpts modifies the loss function in deep learning models to reduce the dependencies between error distributions and racial and ethnic labels. Our results show that DemOpts improves both the hard and soft error parity of COVID-19 county case predictions when compared with state-of-the-art debiasing methods.

Funding

This research is based on work supported by the National Science Foundation Award NO 2210572 (Bringing Transparency and Interpretability to Bias Mitigation Approaches in Place-based Mobility-centric Prediction Models for Decision Making in High-Stakes Settings) to the University of Maryland. There was no direct involvement of the funding institution in the design, analysis, and interpretation of the results.

Data Availability

We use open source data: (1) SafeGraph mobility, (2) New York Times COVID-19 case count, and demographic information from the American Communities Survey (ACS).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details on the differentiability and regression analysis of fairness method errors.
[PDF File, 257 KB - ojphi_v18i1e78235_app1.pdf]

References

1. COVID-19 forecasting and mathematical modeling. Centers for Disease Control and Prevention (CDC). 2023. URL: https://www.cdc.gov/coronavirus/2019-ncov/science/forecasting/forecasting-math-modeling.html [accessed 2025-12-27]
2. COVID-19 Forecast Hub. 2020. URL: https://covid19forecasthub.org/ [accessed 2025-12-27]
3. Zou D, Wang L, Xu P, Chen J, Zhang W, Gu Q. Epidemic model guided machine learning for COVID-19 forecasts in the United States. medRxiv. Preprint posted online on May 24, 2020. [doi: 10.1101/2020.05.24.20111989]
4. Pei S, Shaman J. Initial simulation of SARS-CoV2 spread and intervention effects in the continental US. medRxiv. Preprint posted online on Mar 21, 2020. [doi: 10.1101/2020.03.21.20040303]
5. Chiang WH, Liu X, Mohler G. Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. medRxiv. Preprint posted online on Jun 6, 2020. [doi: 10.1101/2020.06.06.20124149]

XSL•FO
RenderX

6. Galasso J, Cao DM, Hochberg R. A random forest model for forecasting regional COVID-19 cases utilizing reproduction number estimates and demographic data. Chaos Solitons Fractals 2022 Mar;156:111779. [doi: 10.1016/j.chaos.2021.111779] [Medline: 35013654]
7. Arik S, Li CL, Yoon J, Sinha R, Epshteyn A, Le L, et al. Interpretable sequence learning for COVID-19 forecasting. arXiv. Preprint posted online on Aug 3, 2020. [doi: 10.48550/arXiv.2008.00646]
8. Zhang-James Y, Hess J, Salkin A, et al. A seq2seq model to forecast the COVID-19 cases, deaths and reproductive R numbers in US counties. Res Sq 2021 Apr 26:rs.3.rs-456641. [doi: 10.21203/rs.3.rs-456641/v1] [Medline: 34013251]
9. Le M, Ibrahim M, Sagun L, Lacroix T, Nickel M. Neural relational autoregression for high-resolution COVID-19 forecasting. 2021 Presented at: epiDAMIK 2021: 4th International Workshop on Epidemiology meets Data Mining and Knowledge Discovery, ACM; Aug 14-18, 2021 URL: https://epidamik.github.io/2021/papers/epiDAMIK_40_paper_12.pdf [accessed 2026-01-14]
10. Lucas B, Vahedi B, Karimzadeh M. A spatiotemporal machine learning approach to forecasting COVID-19 incidence at the county level in the USA. Int J Data Sci Anal 2023;15(3):247-266. [doi: 10.1007/s41060-021-00295-9] [Medline: 35071733]
11. Angulo FJ, Finelli L, Swerdlow DL. Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. JAMA Netw Open 2021 Jan 4;4(1):e2033706. [doi: 10.1001/jamanetworkopen.2020.33706] [Medline: 33399860]
12. Times TNY. Coronavirus (covid-19) data in the United States. GitHub. 2020. URL: https://github.com/nytimes/covid-19-data [accessed 2025-12-27]
13. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
14. Smittenaar P, Stewart N, Sutermaster S, et al. A COVID-19 community vulnerability index to drive precision policy in the US. medRxiv. Preprint posted online on May 19, 2021. [doi: 10.1101/2021.05.19.21257455]
15. COVID-19 community mobility reports. : Google; 2022 URL: https://www.google.com/covid19/mobility/ [accessed 2025-12-27]
16. Mobility trends reports. : Apple; 2022 URL: https://covid19.apple.com/mobility [accessed 2025-12-27]
17. Gursoy F, Kakadiaris IA. Error parity fairness: testing for group fairness in regression tasks. arXiv. Preprint posted online on 2022 URL: https://doi.org/10.48550/arXiv.2208.08279 [accessed 2026-01-16]
18. Erfani A, Frias-Martinez V. A fairness assessment of mobility-based COVID-19 case prediction models. PLoS ONE 2023;18(10):e0292090. [doi: 10.1371/journal.pone.0292090] [Medline: 37851681]
19. Del Rios M, Puente S, Vergara-Rodriguez P, Sugrue N. Latino invisibility in the pandemic. AMA J Ethics 2022 Apr 1;24(4):E289-E295. [doi: 10.1001/amajethics.2022.289] [Medline: 35405055]
20. Coston A, Guha N, Ouyang D, Lu L, Chouldechova A, Ho DE. Leveraging administrative data for bias audits: assessing disparate coverage with mobility data for COVID-19 policy. Presented at: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency [Internet]; Mar 3-10, 2021; New York, NY p. 173-184. [doi: 10.1145/3442188.3445881]
21. Meraihi Y, Gabis AB, Mirjalili S, Ramdane-Cherif A, Alsaadi FE. Machine learning–based research for COVID-19 detection, diagnosis, and prediction: a survey. SN Comput Sci 2022;3(4):286. [doi: 10.1007/s42979-022-01184-z] [Medline: 35578678]
22. Lim B, Arık S, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. Int J Forecast 2021 Oct;37(4):1748-1764. [doi: 10.1016/j.ijforecast.2021.03.012]
23. 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]
24. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y, et al. Graph attention networks stat. Stat (Int Stat Inst) 2017;1050(20):10-48550. [doi: 10.48550/arXiv.1710.10903]
25. Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. Presented at: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; Jul 13-19, 2018. [doi: 10.24963/ijcai.2018/505]
26. Duan L, Hu T, Cheng E, Zhu J, Gao C. Deep convolutional neural networks for spatiotemporal crime prediction. Presented at: Proceedings of the International Conference on Information and Knowledge Engineering (IKE); 2017 p. 61-67 URL: https://www.semanticscholar.org/paper/Deep-Convolutional-Neural-Networks-for-Crime-Duan-Hu/407b84eb7ecfea0557b1a3e38ff0767b9bf25622 [accessed 2026-01-16]
27. Li S, Jin X, Xuan Y, Zhou X, Chen W, Wang YX, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In: Wallach H, Larochelle H, Beygelzimer A, d' ABF, Fox E, Garnett R, editors. Presented at: Proceedings of the 33rd International Conference on Neural Information Processing Systems; Dec 8-14, 2019 URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/6775a0635c302542da2c32aa19d86be0-Paper.pdf [accessed 2025-12-27]
28. Zhou H, Zhang S, Peng J, et al. Informer: beyond efficient transformer for long sequence time-series forecasting. arXiv. Preprint posted online on Dec 14, 2020 URL: https://arxiv.org/abs/2012.07436 [accessed 2025-12-27]

XSL•FO
RenderX

29. Wu H, Xu J, Wang J, Long M. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. arXiv. Preprint posted online on Jun 24, 2021 URL: https://arxiv.org/abs/2106.13008 [accessed 2025-12-27]
30. Zhou T, Ma Z, Wen Q, Wang X, Sun L, Jin R. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting. arXiv. Preprint posted online on Jan 30, 2022 URL: https://arxiv.org/abs/2201.12740 [accessed 2025-12-27]
31. Liu S, Yu H, Liao C, Li J, Lin W, Liu AX, et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting. Presented at: International Conference on Learning Representations (ICLR); Apr 25-29, 2022 URL: https://openreview.net/forum?id=0EXmFzUn5I [accessed 2025-12-27]
32. Nie YH, Sinthong P, Kalagnanam J, Nguyen N. A time series is worth 64 words: long-term forecasting with transformers. arXiv. Preprint posted online on Mar 5, 2023. [doi: 10.48550/arXiv.2211.14730]
33. Vieira MR, Frías-Martínez E, Bakalov P, Frías-Martínez V, Tsotras VJ. Querying spatio-temporal patterns in mobile phone-call databases. Presented at: 2010 Eleventh International Conference on Mobile Data Management; May 23-26, 2010. [doi: 10.1109/MDM.2010.24]
34. Hernandez M, Hong L, Frias-Martinez V, Whitby A, Frias-Martinez E. Estimating poverty using cell phone data: evidence from guatemala. : World Bank Policy Research Working Paper; 2017 URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/122541487082260120/estimating-poverty-using-cell-phone-data-evidence-from-guatemala [accessed 2026-01-16]
35. Frias-Martinez V, Virseda J. Cell phone analytics: Scaling human behavior studies into the millions. Information Technologies & International Development 2013;9(2):35 [FREE Full text]
36. Rubio A, Frias-Martinez V, Frias-Martinez E, Oliver N. Human mobility in advanced and developing economies: a comparative analysis. Presented at: 2010 AAAI Spring Symposium Series; Mar 22-24, 2010 URL: http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1095 [accessed 2026-01-16]
37. Wu J, Frias-Martinez E, Frias-Martinez V. Spatial sensitivity analysis for urban hotspots using cell phone traces. Environ Plan B: Urban Anal City Sci 2021 Nov;48(9):2623-2639. [doi: 10.1177/2399808320985843]
38. Wu J, Abrar SM, Awasthi N, Frias-Martinez E, Frias-Martinez V. Enhancing short-term crime prediction with human mobility flows and deep learning architectures. EPJ Data Sci 2022;11(1):53. [doi: 10.1140/epjds/s13688-022-00366-2] [Medline: 36406335]
39. Wu J, Abrar SM, Awasthi N, Frías-Martínez V. Auditing the fairness of place-based crime prediction models implemented with deep learning approaches. Comput Environ Urban Syst 2023 Jun;102:101967. [doi: 10.1016/j.compenvurbsys.2023.101967]
40. Wesolowski A, Eagle N, Tatem AJ, et al. Quantifying the impact of human mobility on malaria. Science 2012 Oct 12;338(6104):267-270. [doi: 10.1126/science.1223467] [Medline: 23066082]
41. Hong L, Fu C, Torrens P, Frias-Martinez V. Understanding citizens' and local governments' digital communications during natural disasters. WebSci '17: Proceedings of the 2017 ACM on Web Science Conference 2017:141-150. [doi: 10.1145/3091478.3091502]
42. Isaacman S, Frias-Martinez V, Frias-Martinez E. Modeling human migration patterns during drought conditions in La Guajira, Colombia. COMPASS '18: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies 2018:1-9. [doi: 10.1145/3209811.3209861]
43. Bengtsson L, Gaudart J, Lu X, et al. Using mobile phone data to predict the spatial spread of cholera. Sci Rep 2015 Mar 9;5(1):8923. [doi: 10.1038/srep08923] [Medline: 25747871]
44. Ghurye J, Krings G, Frias-Martinez V. A framework to model human behavior at large scale during natural disasters. Presented at: 2016 17th IEEE International Conference on Mobile Data Management (MDM); Jun 13-16, 2016; Porto, Portugal p. 18-27. [doi: 10.1109/MDM.2016.17]
45. Hong L, Frias-Martinez V. Modeling and predicting evacuation flows during hurricane Irma. EPJ Data Sci 2020 Dec;9(1):29. [doi: 10.1140/epjds/s13688-020-00247-6]
46. Frias-Martinez V, Virseda J, Frias-Martinez E. Socio-economic levels and human mobility. Qual Meets Quant Workshop-QMQ 2010:1-6 [FREE Full text]
47. Fu C, McKenzie G, Frias-Martinez V, Stewart K. Identifying spatiotemporal urban activities through linguistic signatures. Comput Environ Urban Syst 2018 Nov;72:25-37. [doi: 10.1016/j.compenvurbsys.2018.07.003]
48. Frias-Martinez V, Virseda J, Gomero A. Mobilizing education: evaluation of a mobile learning tool in a low-income school. MobileHCI '12: 14th International Conference on Human Computer Interaction with Mobile Devices and Services 2012:441-450. [doi: 10.1145/2371574.2371642]
49. Hong L, Frias-Martinez E, Frias-Martinez V. Topic Models to Infer Socio-Economic Maps. AAAI 2016;30(1):3835-3841. [doi: 10.1609/aaai.v30i1.9892]
50. Frias-Martinez V, Soto V, Virseda J, Frias-Martinez E. Computing cost-effective census maps from cell phone traces. Department of Computer Science, Colombia University. 2012. URL: http://www.cs.columbia.edu/~vsoto/files/purba12.pdf [accessed 2026-01-19]
51. Kang W, Oshan T, Wolf LJ, et al. A roundtable discussion: defining urban data science. Environ Plan B Urban Anal City Sci 2019 Nov;46(9):1756-1768. [doi: 10.1177/2399808319882826]

XSL•FO
RenderX

52. Hong L, Wu J, Frias-Martinez E, Villarreal A, Frias-Martinez V. Characterization of internal migrant behavior in the immediate post-migration period using cell phone traces. Presented at: Proceedings of the Tenth International Conference on Information and Communication Technologies and Development; Jan 4-7, 2019; Ahmedabad, India p. 1-12. [doi: 10.1145/3287098.3287119]
53. Frias-Martinez V, Sloate E, Manglunia H, Wu J. Causal effect of low-income areas on shared dockless e-scooter use. Transp Res D Trans Environ 2021 Nov;100:103038. [doi: 10.1016/j.trd.2021.103038]
54. Badr HS, Gardner LM. Limitations of using mobile phone data to model COVID-19 transmission in the USA. Lancet Infect Dis 2021 May;21(5):e113. [doi: 10.1016/S1473-3099(20)30861-6] [Medline: 33152270]
55. Tsai TC, Arik S, Jacobson BH, et al. Algorithmic fairness in pandemic forecasting: lessons from COVID-19. NPJ Digit Med 2022 May 10;5(1):59. [doi: 10.1038/s41746-022-00602-z] [Medline: 35538215]
56. Abrar SM, Awasthi N, Smolyak D, Frias-Martinez V. Analysis of performance improvements and bias associated with the use of human mobility data in COVID-19 case prediction models. ACM J Comput Sustain Soc 2023 Dec 31;1(2):1-36. [doi: 10.1145/3616380]
57. Douglas MD, Respress E, Gaglioti AH, et al. Variation in reporting of the race and ethnicity of COVID-19 cases and deaths across US states: April 12, 2020, and November 9, 2020. Am J Public Health 2021 Jun;111(6):1141-1148. [doi: 10.2105/AJPH.2021.306167] [Medline: 33856884]
58. Estiri H, Strasser ZH, Rashidian S, et al. An objective framework for evaluating unrecognized bias in medical AI models predicting COVID-19 outcomes. J Am Med Inform Assoc 2022 Jul 12;29(8):1334-1341. [doi: 10.1093/jamia/ocac070]
59. Gross CP, Essien UR, Pasha S, Gross JR, Wang SY, Nunez-Smith M. Racial and ethnic disparities in population-level Covid-19 mortality. J Gen Intern Med 2020 Oct;35(10):3097-3099. [doi: 10.1007/s11606-020-06081-w] [Medline: 32754782]
60. Agarwal A, Dudík M, Wu ZS. Fair regression: quantitative definitions and reduction-based algorithms. arXiv. Preprint posted online on May 30, 2019 URL: https://arxiv.org/abs/1905.12843 [accessed 2025-12-27]
61. Fitzsimons J, Al Ali A, Osborne M, Roberts S. A general framework for fair regression. Entropy (Basel) 2019 Jul 29;21(8):741. [doi: 10.3390/e21080741] [Medline: 33267455]
62. Brunet ME, Alkalay-Houlihan C, Anderson A, Zemel R. Understanding the origins of bias in word embeddings. arXiv. Preprint posted online on Oct 8, 2018 URL: https://arxiv.org/abs/1810.03611 [accessed 2025-12-27]
63. Calmon FP, Wei D, Ramamurthy KN, Varshney KR. Optimized data pre-processing for discrimination prevention. arXiv. Preprint posted online on Apr 11, 2017 URL: https://arxiv.org/abs/1704.03354 [accessed 2025-12-27]
64. Yan B, Seto S, Apostoloff N. FORML: learning to reweight data for fairness. arXiv. Preprint posted online on Feb 3, 2022 URL: https://arxiv.org/abs/2202.01719 [accessed 2025-12-27]
65. Wang Y, Singh L. Mitigating demographic bias of machine learning models on social media. 2023 Oct 30 Presented at: EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization; Oct 30 to Nov 1, 2023. [doi: 10.1145/3617694.3623244]
66. Yang J, Soltan AAS, Clifton DA. Algorithmic fairness and bias mitigation for clinical machine learning: a new utility for deep reinforcement learning. medRxiv. Preprint posted online on Jun 24, 2022 URL: https://www.medrxiv.org/content/early/2022/06/27/2022.06.24.22276853 [accessed 2025-12-27]
67. Das R, Dooley S. Fairer and more accurate tabular models through NAS. arXiv. Preprint posted online on Oct 18, 2023 URL: https://arxiv.org/abs/2310.12145 [accessed 2025-12-27] [Medline: 37090239]
68. Jagodnik KM, Ray F, Giorgi FM, Lachmann A. Correcting under-reported COVID-19 case numbers: estimating the true scale of the pandemic. medRxiv. Preprint posted online on 2020. [doi: 10.1101/2020.03.14.20036178]
69. Albani V, Loria J, Massad E, Zubelli J. COVID-19 underreporting and its impact on vaccination strategies. BMC Infect Dis 2021 Oct 28;21(1):1111. [doi: 10.1186/s12879-021-06780-7] [Medline: 34711190]
70. Berk R, Heidari H, Jabbari S, Joseph M, Kearns M, Morgenstern J, et al. A convex framework for fair regression. arXiv. Preprint posted online on 2017. [doi: 10.48550/arXiv.1706.02409]
71. Shah A, Bu Y, Lee JK, Das S, Panda R, Sattigeri P, et al. Selective regression under fairness criteria. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, editors. 2022 Presented at: Proceedings of the 39th International Conference on Machine Learning, PMLR p. 19598-19615 URL: https://proceedings.mlr.press/v162/shah22a.html [accessed 2025-12-27]
72. Cramer EY, Huang Y, Wang Y, et al. The United States COVID-19 Forecast Hub dataset. Sci Data 2022 Aug 1;9(1):462. [doi: 10.1038/s41597-022-01517-w] [Medline: 35915104]
73. Kader F, Smith CL. Participatory approaches to addressing missing COVID-19 race and ethnicity data. Int J Environ Res Public Health 2021 Jun 18;18(12):6559. [doi: 10.3390/ijerph18126559] [Medline: 34207130]
74. Blanca MJ, Alarcón R, Arnau J, Bono R, Bendayan R. Non-normal data: is ANOVA still a valid option? Psicothema 2017 Nov;29(4):552-557. [doi: 10.7334/psicothema2016.383] [Medline: 29048317]
75. Zimmerman DW. Comparative power of student t test and Mann-Whitney U test for unequal sample sizes and variances. The Journal of Experimental Education 1987:171-174. [doi: 10.1080/00220973.1987.10806451]
76. Castelnovo A, Crupi R, Greco G, Regoli D, Penco IG, Cosentini AC. A clarification of the nuances in the fairness metrics landscape. Sci Rep 2022 Mar 10;12(1):4209. [doi: 10.1038/s41598-022-07939-1] [Medline: 35273279]

XSL•FO
RenderX

77. Kang Y, Gao S, Liang Y, Li M, Rao J, Kruse J. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Sci Data 2020 Nov 12;7(1):390. [doi: 10.1038/s41597-020-00734-5] [Medline: 33184280]
78. Schlosser F, Sekara V, Brockmann D, Garcia-Herranz M. Biases in human mobility data impact epidemic modeling. arXiv. Preprint posted online on Dec 23, 2021 URL: https://arxiv.org/abs/2112.12521 [accessed 2025-12-27]
79. Mapbox. URL: https://www.mapbox.com/about/maps [accessed 2026-01-16]
80. OpenStreetMap. URL: https://www.openstreetmap.org/about/ [accessed 2026-01-16]
81. Improve this map. URL: https://apps.mapbox.com/feedback/ [accessed 2026-01-16]
82. PyTorch-based forecasting with sktime integration sktime. 2020. URL: https://github.com/sktime/pytorch-forecasting [accessed 2026-01-16]
83. Gatalo O, Tseng K, Hamilton A, Lin G, Klein E, CDC MInD-Healthcare Program. Associations between phone mobility data and COVID-19 cases. Lancet Infect Dis 2021 May;21(5):e111. [doi: 10.1016/S1473-3099(20)30725-8] [Medline: 32946835]
84. Kim JS, Chen J, Talwalkar A. FACT: a diagnostic for group fairness trade-offs. ICML'20: Proceedings of the 37th International Conference on Machine Learning 2020:5264-5274 [FREE Full text]

Abbreviations

AER: Accuracy Equity Ratio
CDC: Centers for Disease Control and Prevention
HSD: honestly significant difference
NYT: New York Times
OD: origin-destination
PBL: pinball loss
TFT: temporal fusion transformer

Edited by E Mensah; submitted 29.May.2025; peer-reviewed by B Amuda, J Song, L Rountree; revised version received 20.Sep.2025; accepted 10.Dec.2025; published 03.Feb.2026.

Please cite as:
Awasthi N, Abrar S, Smolyak D, Frias-Martinez V
Fairness Correction in COVID-19 Predictive Models Using Demographic Optimization: Algorithm Development and Validation Study
Online J Public Health Inform 2026;18:e78235
URL: https://ojphi.jmir.org/2026/1/e78235
doi:10.2196/78235

© Naman Awasthi, Saad Abrar, Daniel Smolyak, Vanessa Frias-Martinez. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 3.Feb.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Data Access, User Engagement, and Interoperability Standards in Public Health Data Portals: Environmental Scan Study

Yogita Chaudhary[1], MPH; Lathika R[1], MPH; Solomon Thirumurugan, MSc; Manikandan Kumaraguru[1], BE; Sharan Murali[2], MD; Rizwan Suliankatchi Abdulkader[2], PhD

[1]Intramural Research Team, ICMR National Institute of Epidemiology, Chennai, TN, India

[2]ICMR National Institute of Epidemiology, 2nd Main Rd, near Ambattur, Karmari Amman Nagar, TNHB Colony, Bhavani Nagar, Ayappakkam, Chennai, TN, India

Corresponding Author:
Sharan Murali, MD
ICMR National Institute of Epidemiology, 2nd Main Rd, near Ambattur, Karmari Amman Nagar, TNHB Colony, Bhavani Nagar, Ayappakkam, Chennai, TN, India

Abstract

Background: Rapid digitalization has led to the generation of vast volumes of health data across diverse sources, including public health surveillance, electronic health records from hospitals, primary care facilities, and telemedicine. However, it often remains fragmented across multiple domains and systems due to siloed data collection practices and a lack of interoperability standards, limiting its potential to generate actionable public health insights.

Objective: In this study, we examined publicly available health data portals and assessed their scope, user engagement features, and implementation of interoperability standards.

Methods: We conducted an environmental scan to profile publicly available health data portals. We identified portals through systematic web searches, expert consultation using the Delphi technique, and snowball sampling from initial sources. We assessed each portal for the type of data hosted, the scope of user engagement, and the implementation of interoperability standards (Systematized Nomenclature of Medicine Clinical Terms [SNOMED CT]) or adopted Fast Healthcare Interoperability Resources (FHIR) as a data exchange standard.

Results: We identified 17 data portals in total, of which 9 (53%) hosted global data and 1 (6%) hosted regional data. Portals hosted data across diverse domains including communicable diseases and noncommunicable diseases, maternal and child health, injuries, and others. Most portals supported basic user interactions such as data querying (n=15, 88%) and downloading (n=15, 88%), but none offered in-portal analytics. While 3 (18%) portals used *International Classification of Diseases* (*ICD*) coding systems, none implemented SNOMED CT or adopted FHIR.

Conclusions: Our study highlighted substantial fragmentation and a lack of interoperability across health data portals, which limits the ability to conduct integrated analysis and obtain comprehensive public health insights. Realizing the full potential of health data will require accessibility and interoperability, intelligent analytics, and seamless integration across domains.

(Online J Public Health Inform 2026;18:e95523) doi:10.2196/95523

KEYWORDS

standards; health information system interoperability; digital health; Systematized Nomenclature of Medicine Clinical Terms; SNOMED CT; health information exchange

Introduction

The emphasis on digital health has grown tremendously worldwide in recent years, driven by advances in information and communication technology (ICT) aimed at improving health outcomes [1]. ICT is increasingly harnessed to promote health, prevent diseases, and improve service delivery. A vast amount of digital health care data are generated across domains such as public health surveillance and disease outbreaks, electronic health records (EHRs) from hospitals and primary care settings, and telemedicine, which are lying fragmented and siloed [2-6]. Fragmented data systems pose a major barrier to fully leveraging existing data sources to generate actionable knowledge that can be translated into practice [7]. There is a lack of adherence to the 4 principles of data management—findability, accessibility, interoperability, and reusability (FAIR)—which limits effective data exchange and its use [8]. Addressing these challenges requires innovative approaches.

Big data solutions, which aim to extract knowledge and insights from large and complex digital data, offer promising opportunities. They can integrate the silos to generate insights for better epidemiology surveillance, research, and efficient

XSL•FO
RenderX

health care delivery [2]. Clinical vocabularies, data standards, and interoperability are an integral part of big data solutions and effective digital health strategies [2,9]. Interoperable systems enable accurate data collection and sharing across countries and organizations without information loss [1,9,10].

Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) are among the top recommended standards to achieve interoperability [10-12]. SNOMED CT is a formally structured, consensually developed clinical terminology that provides standardized representations of clinical and public health concepts, enabling semantic interoperability across systems. SNOMED CT is the most comprehensive, multilingual clinical health care terminology globally, with more than 360,000 concepts. It has three core components: (1) concepts—represent a unique clinical term; (2) relationships—define the association between two concepts; and (3) descriptions—provide human-readable terms to convey the meaning of the clinical term [13]. SNOMED CT enables the consistent encoding of clinical and public health data, helping researchers worldwide better understand regional data and integrate them with other datasets [14,15].

HL7 FHIR, by contrast, is a data exchange standard that defines how health-related information is structured and transmitted between systems. HL7 FHIR standards were developed by the HL7 organization in 2011. It has two primary components: (1) resources—determine the structure and elements of exchangeable data—and (2) APIs—enable interoperability between the applications. It supports various health care domains through resources such as “patient,” “procedure,” and “observation,” making it adaptable for different purposes and contexts [16].

In recent years, an increasing number of countries have adopted HL7 FHIR and SNOMED CT standards to implement interoperability frameworks [10,17,18]. These standards are commonly used for EHRs in the health system [10]. However, their adoption is limited in the public health system [19]. We conducted an environmental scan of publicly accessible digital health data portals to describe the variety of health data hosted and their scope of user engagement. We also evaluated the implementation of HL7 FHIR and SNOMED CT standards in these portals.

Methods

An environmental scan of digital health data portals was conducted between December 2023 and May 2024.

Search Strategy

We used the following methods to profile digital portals that host health databases:

1. Search engines: we used Google as our primary search engine, using keywords such as “online healthcare portals,” “online healthcare data portals,” “digital health portals,” “healthcare data portals,” “health data portals,” “data portals,” “public health,” and “population level data” both in combinations and separately.
2. Delphi technique: we consulted scientists and epidemiologists at the Indian Council of Medical Research (ICMR), National Institute of Epidemiology (NIE), Chennai, to identify health-related data portals of which they were aware.
3. Snowballing from initial sources: we identified additional relevant portals by snowball sampling from the portals identified through methods 1 and 2.

Inclusion Criteria

We included digital portals that store any type of health-related data, are publicly accessible, and are available free of charge. We excluded portals whose primary language was not English.

Operational Definition

We used the following operational definitions throughout our study:

- Data analysis (also referred to as in-portal analysis): data analysis refers to users’ ability to perform computational or statistical operations on raw data within the portal, including deriving new variables and examining relationships between the existing ones.
- Data granularity: data granularity refers to the level at which data are collected, stored, and presented [20]. The granularity of data hosted is categorized into 3 levels: global, country, or regional.
- Download data: downloading data refers to the ability of users to download datasets of interest from the portal in any format for offline use.
- Health data portals: health data portals are web-based interfaces that organize and present health-related data from various sources for discovery and efficient use by policymakers, researchers, journalists, and the public [21].
- Query the data: querying, in the context of database management, refers to the process of requesting specific data and retrieving the information as a table or graph, based on preaggregated, fixed indicators [22].
- Upload data: uploading data refers to the ability of users to contribute their own datasets to the portal for data sharing and collaborative use, as well as to perform computational or statistical operations on the uploaded data within the portal.
- Ontologies: we defined ontology as a formally structured, consensually developed clinical terminologies and coding systems that provide standardized representations of clinical and public health concepts.

Data Extraction and Analysis

We reviewed the included portals and extracted their aims, visions, missions, ownership statuses, data granularity, types of health data hosted, scopes of user engagement, and adoption of SNOMED CT as a clinical terminology and HL7 FHIR as a data exchange standard onto an Excel spreadsheet. We summarized the characteristics of the portals in a table and visualized the types of health data hosted, user engagement scope, and adoption of standards in an infographic. Given the Indian research context of this study, we described the Indian digital health data portals as case studies to provide deeper insights into their capabilities and limitations. From the list of

XSL•FO
RenderX

non-Indian portals, we selected 3 portals that represented a diversity of ownership, including an intergovernmental body, a government agency, and a collaborative nonprofit initiative.

We categorized the health data hosted in the study portals into five domains: (1) communicable disease (CD); (2) noncommunicable disease (NCD); (3) injuries; (4) reproductive, maternal, newborn, child, and adolescent health and nutrition (RMCH+N); and (5) others. We included data that did not fit into the primary categories in the "others" category.

The scope of user engagement with the portal was classified at 4 levels—the ability of a user to query the data, upload data, analyze data, and download data. We summarized the findings on portal ownership, level of data granularity, data domains, scope of user engagement, and data standards as frequencies and proportions.

Ethical Considerations

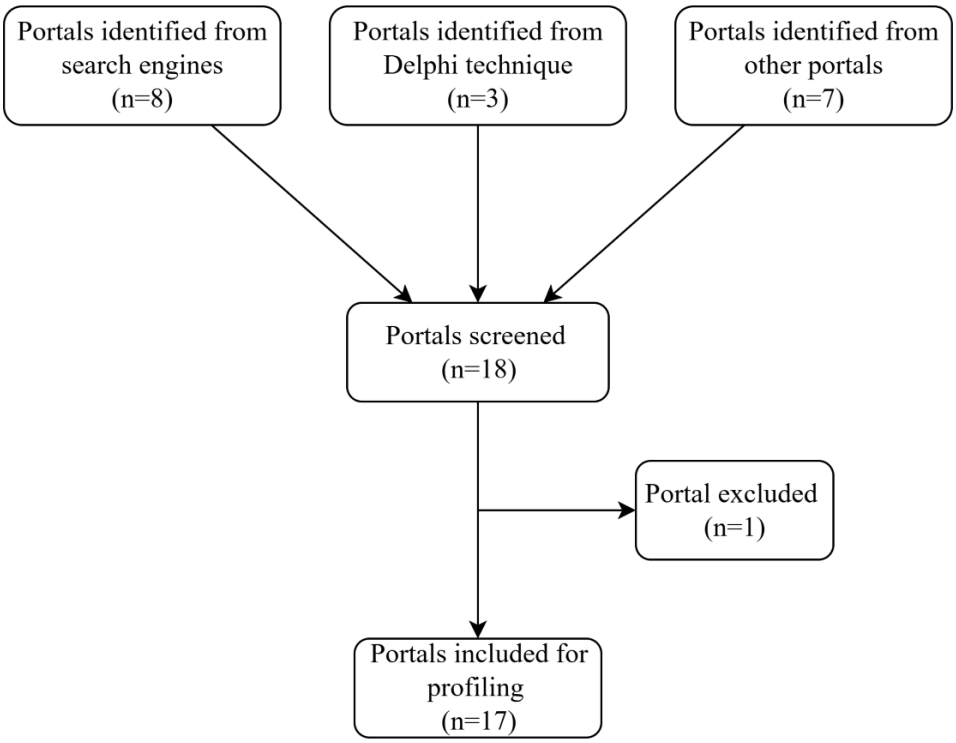
This study is part of a larger ICMR-NIE project titled "ICMR Epihub: A platform for epidemic intelligence and interpretation of infectious disease epidemiology." This study has been approved by the Institutional Human Ethics Committee (IHEC) of ICMR-NIE (NIE/IHEC/202210‐03). Personal information was not included in this study; thus, informed consent was not required.

Results

Overview

A total of 18 portals met the inclusion criteria, of which 1 portal was excluded because its primary language was not English (Figure 1). The overarching objective of these portals was to enhance the accessibility of health care data, thereby improving the understanding of health outcomes for individuals and nations (Table 1).

Figure 1. Health data portals included in the analysis.



XSL•FO
RenderX

Table . Overview of digital health portals: objectives, ownership, and data granularity.

Portal name	Ownership	Data granularity
Centers for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research [22]	Government (US Department of Health and Human Services)	Country (United States)
Gapminder [23]	Nonprofit (Gapminder Foundation)	Global
Health Indicators Warehouse [24]	Government (US Department of Health and Human Services)	Country (United States)
Institute for Health Metrics and Evaluation portal [25]	Academic (IHME[a], University of Washington)	Global
India Policy Insights [26]	Collaborative (Geographic Insights Lab, IIPS[b], and NITI[c] Aayog)	Country (India)
National Data and Analytics Platform [27]	Government (NITI Aayog)	Country (India)
Noncommunicable Disease Risk Factor Collaboration [28]	Collaborative (WHO[d] and academic partners)	Global
National Center for Health Statistics Data Query System [29]	Government (US Department of Health and Human Services)	Country (United States)
Organisation for Economic Co-operation and Development Data Explorer [30]	Intergovernmental (OECD[e])	Country (OECD+selected non-OECD countries)
Our World in Data [31]	Collaborative (Global Change Data Lab and University of Oxford)	Global
Rochester Epidemiology Project data exploration portal [32]	Collaborative (County of Olmsted, Mayo Clinic, Mayo Clinic Health System, Olmsted Medical Center, and Zumbro Valley Health Center)	Regional
State of Global Air [33]	Collaborative (Health Effects Institute and IHME)	Global
The Global Health Observatory [34]	UN[f] agency (WHO)	Global
The Socioeconomic High-Resolution Rural-Urban Geographic [35]	Nonprofit (Development Data Lab)	Country (India)
United Nations Inter-agency Group for Child Mortality Estimation [36]	Collaborative (UNICEF[g], UN DESA[h], WHO, and World Bank)	Global
United Nations Population Division Data Portal [37]	UN agency (UN DESA)	Global
United Nations Sustainable Development Goals Indicators database [38]	UN agency (UN DESA)	Global

[a]IHME: Institute for Health Metrics and Evaluation.
[b]IIPS: International Institute for Population Sciences.
[c]NITI: National Institution for Transforming India.
[d]WHO: World Health Organization.
[e]OECD: Organisation for Economic Co-operation and Development.
[f]UN: United Nations.
[g]UNICEF: United Nations Children's Fund.
[h]UN DESA: United Nations Department of Economic and Social Affairs.

Ownership

Of the 17 health data portals, governments own 4 (24%) portals, nonprofit organizations own 2 (12%) portals, and 6 (35%) portals are managed in collaboration. The remaining 5 (29%) portals are maintained by academic institutions, intergovernmental bodies, or United Nations agencies (Table 1).

Among the 4 government-owned portals, the US Department of Health and Human Services manages 3 (75%) portals, namely, the National Center for Health Statistics Data Query System, Health Indicators Warehouse, and Centers for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research (CDC Wonder) [22,24,29]. In India, the National Institution for Transforming India Aayog manages the National Data and Analytics Platform (NDAP) [27].

XSL•FO
RenderX

Level of Data Granularity

Of the 17 portals, 9 (53%) provided health estimates at the global level, 7 (41%) reported data at the country level, while 1 (6%) reported data at the regional level (Table 1). Among the country-level portals, 3 portals focus exclusively on India-specific data, namely, NDAP, India Policy Insights, and Socioeconomic High-Resolution Rural-Urban Geographic (SHRUG) [26,27,35]. Three portals focus exclusively on US-specific data, and the Organisation for Economic Co-operation and Development (OECD) Data Explorer focuses on OECD and selected non-OECD countries [22,24,29,30].

Domains Covered by the Data

A total of 11 (65%) portals hosted data related to CD, 12 (71%) had data on NCD, 12 (71%) had data on RMCH+N; 8 (47%) portals hosted data related to injuries, and 7 (41%) portals, namely, CDC Wonder, Gapminder, Health Indicator Warehouse, Institute for Health Metrics and Evaluation (IHME) portal, OECD Data Explorer, the Global Health Observatory (GHO), and United Nations Sustainable Development Goals Indicators database, provided extensive data on all 5 categories (Figure 2) [22-25,30,34,38]. The "other" category includes data related to health behavior; health economics and policy; demography and population; environment and pollution; health systems and infrastructure; pharmaceutical markets; and specialized areas such as oral health, antimicrobial resistance, and the sustainable development goals (SDGs).

Figure 2. Insights into data hosted and the scope of user engagement in the portals reviewed. ..CDC Wonder: Centers for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research; IHME: Institute for Health Metrics and Evaluation; NCD: noncommunicable disease; NCHS: National Center for Health Statistics; OECD: Organisation for Economic Co-operation and Development; RMCH+N: reproductive, maternal, newborn, child and adolescent health and nutrition; SDG: sustainable development goal; SHRUG: Socioeconomic High-Resolution Rural-Urban Geographic; UN: United Nations.

	Health Data Hosted					Scope of User Engagement			
Portal name	Communicable diseases	Non Communicable diseases	Injuries	RMCH&N	Others	Data upload option	Query the data	Data analysis option	Data downloadable
Centers for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Gapminder	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
Health Indicators Warehouse	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Institute for Health Metrics and Evaluation portal	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
India Policy Insights	No	No	No	No	Yes	No	Yes	No	No
National Data and Analytics Platform	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Noncommunicable Disease Risk Factor Collaboration	No	Yes	No	No	No	No	Yes	No	Yes
National Center for Health Statistics Data Query System	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Organisation for Economic Co-operation and Development Data Explorer	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Our World in Data	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Rochester Epidemiology Project data exploration portal	No	No	No	No	Yes	No	Yes	No	No
State of Global Air	No	No	No	No	Yes	No	Yes	No	Yes
The Global Health Observatory	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
The Socioeconomic High-Resolution Rural-Urban Geographic	No	No	No	No	Yes	Yes	No	No	Yes
United Nations Inter-agency Group for Child Mortality Estimation	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes
United Nations Population Division Data Portal	No	No	No	Yes	Yes	No	Yes	No	Yes
United Nations Sustainable Development Goals Indicators database	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes

Scope of User Engagement

The scope of user engagement varied across the portals reviewed. Of the 17 portals, 15 (88%) allowed users to query the data directly, except for Gapminder and SHRUG, which did not support data querying [23,35]. Similarly, 15 (88%) portals, except India Policy Insights and the Rochester Epidemiology Project (REP) data exploration portal, supported data downloads [26,32]. The India Policy Insights portal helps visualize queried data as maps; only these maps are downloadable [26]. The REP data exploration portal does not allow users to download the data from their database in any form [32]. Only 1 (6%) portal, the SHRUG, permitted users to upload their data as a contribution to their database [35]. None of the portals had in-portal analytics capabilities (Figure 2).

Data Standards Used

Of the 17 portals, 3 (18%), namely CDC Wonder, IHME, and GHO, stored the data using *International Classification of Diseases* (*ICD*) codes [22,25,34]. While a few portals, notably

XSL•FO
RenderX

the GHO, supported programmatic data access through APIs, none adopted an FHIR for structured data exchange. This pattern was consistent across both Indian and international portals, suggesting that API-based access, when available, primarily serves as a convenience tool for data retrieval rather than as a mechanism for semantic interoperability.

Selected Non-Indian Portals

The GHO, managed by the World Health Organization (WHO), provided access to more than 1000 indicators on priority health topics across its 194 member states, with a vision to promote health, keep the world safe, and serve the vulnerable [34]. It hosts data spanning CDs and NCDs, maternal and child health, health systems, and the SDGs, drawn from household surveys, civil registration systems, and facility-based sources. Users can query and download data across all domains. Notably, the GHO is the only portal reviewed to support standardized programmatic data access via an OData-compliant API, enabling users and software applications to retrieve datasets in a consistent and predictable way. Health indicators are also being transitioned to universally unique identifiers to strengthen long-term interoperability across datasets and partner systems.

CDC Wonder, managed by the US Department of Health and Human Services, provides access to a wide range of public health data for the United States, spanning CDs and NCDs, mortality, natality, and environmental health [22]. It is 1 of 3 portals that encodes data using the *ICD* system, providing a degree of semantic standardization for cause-of-death and disease classification. Users can query and download data, and the portal supports programmatic access through a proprietary XML-based API.

Our World in Data (OWID), a collaborative initiative of the Global Change Data Lab and the University of Oxford, aims to make data on the world's largest problems accessible to a broad audience [31]. It hosts data spanning health, poverty, education, climate, and demography, drawn from international sources including the WHO, the World Bank, and academic institutions. Users can query, visualize, and download data. More recently, the portal introduced a Chart Data API enabling access to datasets. OWID aggregated data from diverse original sources with differing coding conventions. There is no harmonization of health concepts across datasets, which limits their utility for cross-dataset analytical work.

Indian Portals

NDAP is developed with a vision to make data generated by the government accessible to the public [27]. All survey data spanning 31 sectors and 53 ministries are available as aggregates for the fixed indicators. All datasets are standardized to a common schema and mapped to a unified geographical framework using the Ministry of Panchayati Raj Local Government Directory Code, enabling users to merge datasets across sectors and sources for cross-sectoral analysis. This portal enables users to create queries based on predefined indicators for each dataset and download the data and results. Notably, NDAP represents one of the more deliberate efforts among the reviewed portals to address interoperability not through health-specific standards such as the FHIR or SNOMED CT but through a common administrative schema that enables linkage across government datasets.

The India Policy Insights portal informs evidence-based health policy decisions through its online geo-visual data portal [26]. The portal aggregated data from various sources, including population censuses, health surveys, the COVID-19 vaccination dashboard (COWIN), Datameet, assembly constituencies, and parliamentary constituencies [39]. It tracks the impact of health policies through an interactive dashboard through various use cases such as the National Family Health Survey Policy Tracker for Parliamentary Constituencies, India COVID-19 Vaccine Tracker, and Visualizing Child Undernutrition across Assembly Constituencies in India. Users can only interact with the data through these fixed dashboards, and there is no option to download the data. The portal does not adopt any recognized data exchange standard, and its fixed dashboard structure limits user engagement to passive visualization rather than active data exploration.

The SHRUG is a research collaboration portal that facilitates data upload and download and was developed by the Development Data Lab, a nonprofit organization [35]. This portal harnesses data from various sources, including population census, socioeconomic caste census, and geospatial data for India. SHRUG also hosts data from special projects, such as Mission Antyodaya Village Facilities and Facebook Wealth and Population statistics. It has generated specific use cases, such as optimal locations for decentralized health care delivery, changes in population characteristics with proximity to towns, and dynamic changes in socioeconomic class structure in India. A distinctive feature of SHRUG is its use of a consistent set of universal geographic identifiers, the shrid, which links all datasets at the village and town level, enabling researchers to merge data from disparate sources without additional harmonization effort. However, similar to the other portals reviewed, SHRUG does not adopt health-specific interoperability standards and does not allow the user to query the data or perform any in-portal analysis.

Discussion

Principal Findings

We reviewed the publicly available health data portals to profile them and identify the scope of user engagement and the use of standard ontologies. We identified 17 portals that enhanced access to health data across domains and geographies. These portals have democratized access to health-related information and empowered users, ranging from policymakers to journalists, to engage with data and insights more readily. However, they lacked capabilities for dynamic user-driven analysis and real-time insights for alerts or clinical decision support. We also identified that none of these platforms used interoperable standards such as SNOMED CT and HL7 FHIR.

Our study identified that health care datasets are publicly available for download and use. We believe this has significantly enhanced access to health-related information, empowering policymakers and researchers to engage with health data more readily and use it to inform their work. Although the public is

XSL•FO
RenderX

unlikely to query or download raw data directly, they typically rely on summary dashboards, visualizations, and narrative reports to interpret health information. The availability of raw, quarriable data nonetheless remains valuable for public health practitioners and other technical users within the health domain, who are better positioned to derive actionable insights from it. Overall, these portals have democratized data access and facilitated transparency and informed decision-making in public health [40,41]. However, we identified that the capacity for users to perform in-depth analyses within these portals remains limited. If these portals do not provide tools to analyze the data or generate insights directly, the responsibility of deriving meaning from the data falls back onto the user and their skills. This can hinder the real-time translation of the data into actionable public health interventions.

We identified that the health data portals are characterized by fragmentation, where disparate datasets are scattered across various portals, each serving distinct purposes. This disjointed approach hampers the ability to perform integrated analyses essential for comprehensive public health insights [42]. For instance, although rainfall data may be available through meteorological departments and malaria incidence data through health surveillance systems, the lack of interoperability between these datasets prevents researchers from effectively correlating climatic factors with disease outbreaks. This disconnect restricts the public health community's capacity to conduct predictive analyses or design targeted interventions [43].

We identified that none of the portals used interoperable standards such as the SNOMED CT and HL7 FHIR. The literature from low- and middle-income countries also highlighted similar challenges, where health data systems operate in silos due to the absence of common metadata schemas, harmonized ontologies, or standardized data-sharing protocols [42,44]. During the COVID-19 pandemic, data standards and infrastructure made it difficult to harness the power of AI and predictive data science to improve clinical care for patients with COVID-19 [11]. The COVID-19 pandemic highlighted the necessity of globally aligned standardization of health care data to respond to future public health crises [11,42]. Without integration, data remain underused, and opportunities for system-wide learning are lost. Interoperable systems can enhance routine and emergency public health activities, such as surveillance and outbreak management, by enabling faster responses and earlier interventions [6]. Thus, there is a need to develop portals that aggregate datasets from diverse sectors using data exchange standards such as HL7 FHIR.

Health data portals should also support layered analysis across domains such as health, climate, demographics, and socioeconomics, empowering stakeholders to act on evidence in real time [42,45]. This will help democratize access to data and facilitate their use for evidence-based action in the field. We must leverage the current digital era and the evolving capabilities of AI to harness the true power of health data. When paired with interoperable frameworks such as the SNOMED CT and HL7 FHIR, they can transform unstructured data into standardized formats, enabling predictive analytics and intelligent decision support systems [46]. Leveraging AI in this way not only supports scalable public health surveillance but also accelerates the development of responsive, data-driven health systems suited for both routine and emergency settings.

Although we performed this analysis in the timely context of postpandemic digital health reforms, our study had some limitations too. First, the search strategy used in this study, which relied on web searches, expert consultation, and reference snowballing, was not exhaustive and may have resulted in the omission of eligible portals, particularly those from non–English-speaking countries or those less prominently indexed online. Second, the assessment of user engagement relied on interface reviews and publicly available documentation. We have not performed any formal testing on or collected empirical user feedback from these portals. This may limit the generalizability of conclusions on user engagement.

Third, the capabilities of these portals were not scored using a standardized appraisal framework, which may introduce some subjectivity in comparisons. Fourth, given the dynamic nature of digital health portals, the findings represent a snapshot in time and may not reflect recent updates or feature enhancements. Finally, the study focused solely on publicly accessible portals and did not include closed systems or internal dashboards used within ministries or large health programs, which may offer richer analytical capabilities. It is important to note that such systems are intentionally restricted from public access, as the depth of data and analytical functionality they provide is neither required nor appropriate for the general public. These internal systems might host more granular data and support more advanced analytical capabilities than those observed in the publicly accessible portals reviewed here. Consequently, the findings and conclusions of this study should be interpreted strictly within the context of publicly accessible health data infrastructure and should not be generalized to the broader ecosystem of health information systems.

Conclusions

Our study highlights the need for health data portals to move beyond data access toward interoperability and richer in-portal analytical capabilities. Addressing fragmentation and limited user functionality is essential to support actionable, equity-focused insights. Future health data portals should adopt standard ontologies and analytical tools to empower diverse users across sectors.

Acknowledgments

All authors declared that they had insufficient funding to support open-access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

XSL•FO
RenderX

We would like to acknowledge the guidance of Dr Prabhdeep Kaur, Professor and Chair, Isaac Centre for Public Health, Indian Institute of Science, and her team in preparing this manuscript. We also acknowledge the support from Professor Doctor Inder Gopal, Mr Rakshit Ramesh, Mr Novoneel Chakraborty, and Mr Amarthya Ravi in getting this work published.

Funding

This project was funded by the Pradhan Mantri Ayushman Bharat Health Infrastructure Mission (grant VIR/29/2022/ECD-1). The funder had no role in the study design, data collection, data analysis, data interpretation, manuscript preparation, or the decision to submit the manuscript for publication. We also acknowledge the funding and technical support from the Isaac Centre for Public Health and the Centre of Data for Public Good, hosted at the Indian Institute of Science, Bengaluru.

Data Availability

Data is presented in the main manuscript.

Authors' Contributions

YC undertook data curation, formal analysis, investigation, methodology, and writing—original draft. LR contributed to data curation and project administration. ST was responsible for visualization. MK contributed to project administration. SM contributed to conceptualization, funding acquisition, methodology, supervision, validation, and writing—review and editing. RSA contributed to funding acquisition, resources, and supervision. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Mamuye AL, Yilma TM, Abdulwahab A, et al. Health information exchange policy and standards for digital health systems in Africa: a systematic review. PLOS Digit Health 2022 Oct;1(10):e0000118. [doi: 10.1371/journal.pdig.0000118] [Medline: 36812615]
2. Fernandes L, O'Connor M, Weaver V. Big data, bigger outcomes: healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis. J AHIMA 2012 Oct;83(10):38-43. [Medline: 23061351]
3. Kabaso B, Korpela M. Architecture for health information systems interoperability in Africa. J Health Inform Afr 2013;1. [doi: 10.12856/JHIA-2013-v1-i1-59]
4. Olu O, Muneene D, Bataringaya JE, et al. How can digital health technologies contribute to sustainable attainment of universal health coverage in Africa? A perspective. Front Public Health 2019;7:341. [doi: 10.3389/fpubh.2019.00341] [Medline: 31803706]
5. Stroetmann KA. Digital health ecosystem for African countries: a guide for public and private actors for establishing holistic digital health ecosystem in Africa. : Federal Ministry for Economic Cooperation and Development; 2019 URL: https://israel-trade.net/wp-content/uploads/2019/03/Materilie345_digital_health_africa.pdf [accessed 2026-08-29]
6. Williams F, Oke A, Zachary I. Public health delivery in the information age: the role of informatics and technology. Perspect Public Health 2019 Sep;139(5):236-254. [doi: 10.1177/1757913918802308] [Medline: 30758258]
7. Vorisek CN, Lehne M, Klopfenstein SA, et al. Fast Healthcare Interoperability Resources (FHIR) for interoperability in health research: systematic review. JMIR Med Inform 2022 Jul 19;10(7):e35724. [doi: 10.2196/35724] [Medline: 35852842]
8. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3:160018. [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
9. National eHealth strategy toolkit. World Health Organization. 2012. URL: https://www.who.int/publications/i/item/national-ehealth-strategy-toolkit [accessed 2026-08-29]
10. Chatterjee A, Pahari N, Prinz A. HL7 FHIR with SNOMED-CT to achieve semantic and structural interoperability in personal health data: a proof-of-concept study. Sensors (Basel) 2022 May 15;22(10):3756. [doi: 10.3390/s22103756] [Medline: 35632165]
11. Gong M, Jiao Y, Gong Y, Liu L. Data standards and standardization: the shortest plank of bucket for the COVID-19 containment. Lancet Reg Health West Pac 2022 Dec;29:100565. [doi: 10.1016/j.lanwpc.2022.100565] [Medline: 35971388]
12. Kwon A, Lee HY, Shin SY, et al. Current health data standardization project and future directions to ensure interoperability in Korea. Healthc Inform Res 2024 Apr;30(2):93-102. [doi: 10.4258/hir.2024.30.2.93] [Medline: 38755100]
13. What is SNOMED CT? SNOMED International. URL: https://www.snomed.org/what-is-snomed-ct [accessed 2024-12-16]
14. Reimer AP, Milinovich A. Using UMLS for electronic health data standardization and database design. J Am Med Inform Assoc 2020 Oct 1;27(10):1520-1528. [doi: 10.1093/jamia/ocaa176] [Medline: 32940707]
15. Wardle M, Spencer A. Implementation of SNOMED CT in an online clinical database. Future Healthc J 2017 Jun;4(2):126-130. [doi: 10.7861/futurehosp.4-2-126] [Medline: 31098449]

XSL•FO
RenderX

16. FHIR overview - architects. Health Level Seven Fast Healthcare Interoperability Resources. URL: https://www.hl7.org/fhir/overview-arch.html [accessed 2024-12-16]
17. Slawomirski L, Lindner L, de Bienassis K, et al. Progress on implementing and using electronic health record systems: developments in OECD countries as of 2021. Organisation for Economic Co-operation and Development. 2023. URL: https://www.oecd.org/en/publications/progress-on-implementing-and-using-electronic-health-record-systems_4f4ce846-en.html [accessed 2026-08-29]
18. SNOMED International members. SNOMED International. URL: https://www.snomed.org/members [accessed 2024-12-16]
19. Hufstedler H, Roell Y, Peña A, et al. Navigating data standards in public health: a brief report from a data-standards meeting. J Glob Health 2024 Apr 5;14:03024. [doi: 10.7189/jogh.14.03024] [Medline: 38577801]
20. Data granularity. Community Commons. URL: https://www.communitycommons.org/entities/69c480a4-7e94-463e-b37d-97f9545254e9 [accessed 2025-02-20]
21. Comprehensive definition: data portal. Huwise. URL: https://www.huwise.com/en/glossary/data-portal/ [accessed 2026-06-17]
22. CDC WONDER. URL: https://wonder.cdc.gov/ [accessed 2025-02-20]
23. Gapminder. URL: https://www.gapminder.org/ [accessed 2025-02-20]
24. HealthData.gov. URL: https://www.healthdata.gov/ [accessed 2025-02-20]
25. Institute for Health Metrics and Evaluation. URL: https://www.healthdata.org/ [accessed 2025-02-20]
26. India Policy Insights. Geographic Insights. URL: https://geographicinsights.iq.harvard.edu/india [accessed 2025-02-20]
27. National Data & Analytics Platform. URL: https://ndap.niti.gov.in/ [accessed 2025-02-20]
28. NCD-RisC (NCD Risk Factor Collaboration). GitHub. URL: https://github.com/NCD-RisC [accessed 2026-09-04]
29. NCHS data query system. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/dqs/index.html [accessed 2025-06-04]
30. OECD data explorer. Organisation for Economic Co-operation and Development. URL: https://data-explorer.oecd.org/ [accessed 2025-02-20]
31. Our World in Data. URL: https://ourworldindata.org/ [accessed 2026-08-29]
32. St Sauver JL, Grossardt BR, Finney Rutten LJ, et al. Rochester Epidemiology Project Data Exploration Portal. Prev Chronic Dis 2018 Apr 12;15:E42. [doi: 10.5888/pcd15.170242] [Medline: 29654640]
33. State of Global Air. URL: https://www.stateofglobalair.org/ [accessed 2025-06-04]
34. Global Health Observatory. World Health Organization. URL: https://www.who.int/data/gho [accessed 2025-02-20]
35. Development Data Lab. URL: https://www.devdatalab.org/shrug [accessed 2025-02-20]
36. CME info - child mortality estimates. UN Inter-agency Group for Child Mortality Estimation. URL: https://childmortality.org/ [accessed 2025-02-20]
37. World Population Prospects 2024. United Nations. URL: https://population.un.org/wpp/ [accessed 2025-02-20]
38. UNSDG. URL: https://unstats.un.org/sdgs/dataportal [accessed 2025-02-20]
39. Data{Meet}. URL: https://datameet.org/ [accessed 2026-06-25]
40. Prathima D. Why is data analysis important for public health? Xebia. 2021. URL: https://xebia.com/blog/data-analysis-important-public-health/ [accessed 2025-07-09]
41. Public health data strategy. : Centers for Disease Control and Prevention; 2023 URL: https://www.cdc.gov/public-health-data-strategy/media/pdfs/2024/06/phds-lookback-report-508.pdf [accessed 2026-08-29]
42. Gardner L, Ratcliff J, Dong E, Katz A. A need for open public data standards and sharing in light of COVID-19. Lancet Infect Dis 2021 Apr;21(4):e80. [doi: 10.1016/S1473-3099(20)30635-6] [Medline: 32791042]
43. Mitsakakis K, Hin S, Müller P, et al. Converging human and malaria vector diagnostics with data management towards an integrated holistic One Health approach. Int J Environ Res Public Health 2018 Feb 3;15(2):259. [doi: 10.3390/ijerph15020259] [Medline: 29401670]
44. Rajendran EG, Mohd Hairi F, Krishna Supramaniam R, T Mohd TA. Precision public health, the key for future outbreak management: a scoping review. Digit Health 2024;10:20552076241256877. [doi: 10.1177/20552076241256877] [Medline: 39139190]
45. Ayushman Bharat Digital Mission. URL: https://abdm.gov.in/ [accessed 2025-07-09]
46. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 4;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]

Abbreviations

CD: communicable disease
CDC Wonder: Centers for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research
EHR: electronic health record
FAIR: findability, accessibility, interoperability, and reusability
FHIR: Fast Healthcare Interoperability Resources
GHO: Global Health Observatory
HL7: Health Level 7

XSL•FO
RenderX

ICD*: *International Classification of Diseases
ICMR: Indian Council of Medical Research
ICT: information and communication technology
IHEC: Institutional Human Ethics Committee
IHME: Institute for Health Metrics and Evaluation
NCD: noncommunicable disease
NDAP: National Data and Analytics Platform
NIE: National Institute of Epidemiology
OECD: Organisation for Economic Co-operation and Development
OWID: Our World in Data
REP: Rochester Epidemiology Project
RMCH+N: reproductive, maternal, newborn, child and adolescent health and nutrition
SDG: sustainable development goal
SHRUG: Socioeconomic High-Resolution Rural-Urban Geographic
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms
WHO: World Health Organization

Edited by E Mensah; submitted 17.Mar.2026; peer-reviewed by D Lee, K Triep; revised version received 26.Jun.2026; accepted 23.Jul.2026; published 07.Oct.2026.

Please cite as:
Chaudhary Y, R L, Thirumurugan S, Kumaraguru M, Murali S, Abdulkader RS
Data Access, User Engagement, and Interoperability Standards in Public Health Data Portals: Environmental Scan Study
Online J Public Health Inform 2026;18:e95523
URL: https://ojphi.jmir.org/2026/1/e95523
doi:10.2196/95523

© Yogita Chaudhary, Lathika R, Solomon Thirumurugan, Manikandan Kumaraguru, Sharan Murali, Rizwan Suliankatchi Abdulkader. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 7.Oct.2026. 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 Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX