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

Viewpoint

Review

Original Papers

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 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.

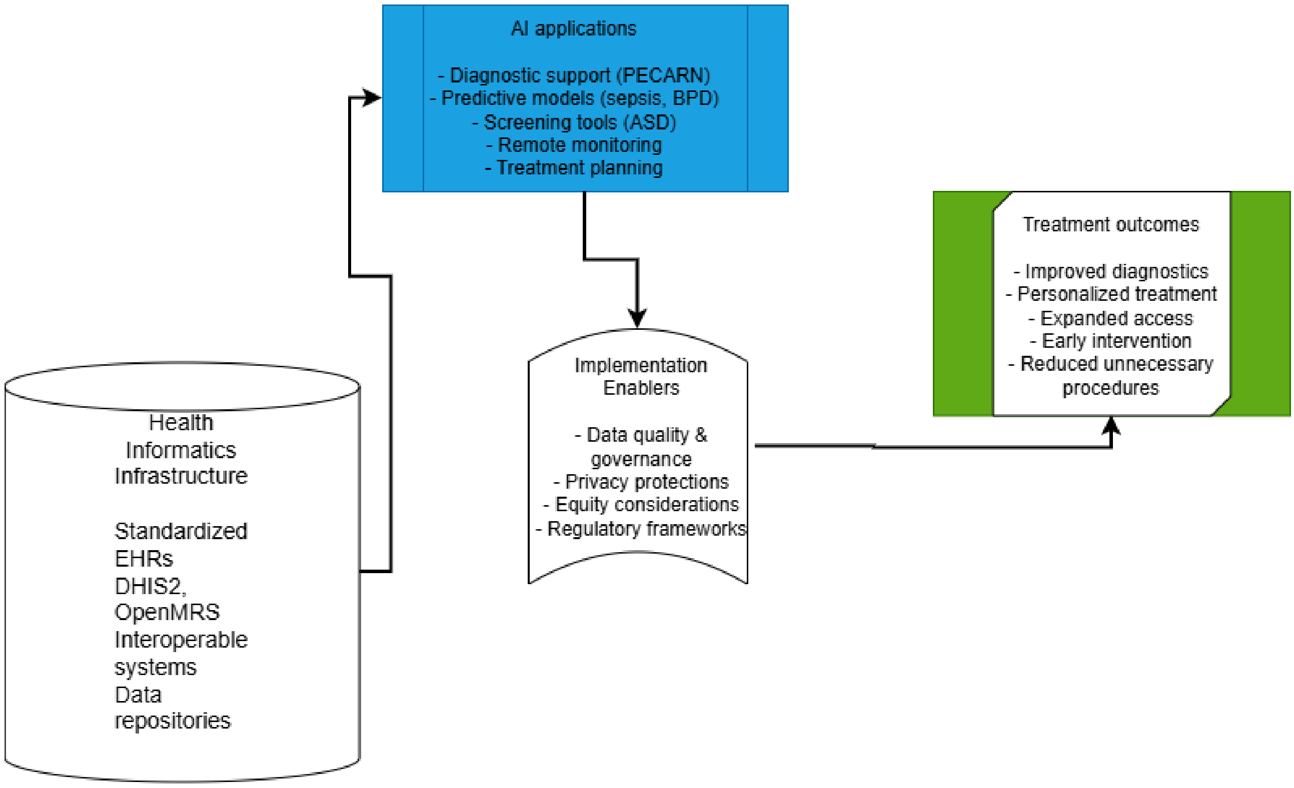
Potential benefits

- 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

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 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 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 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.

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]

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]

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

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.

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.

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.

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].

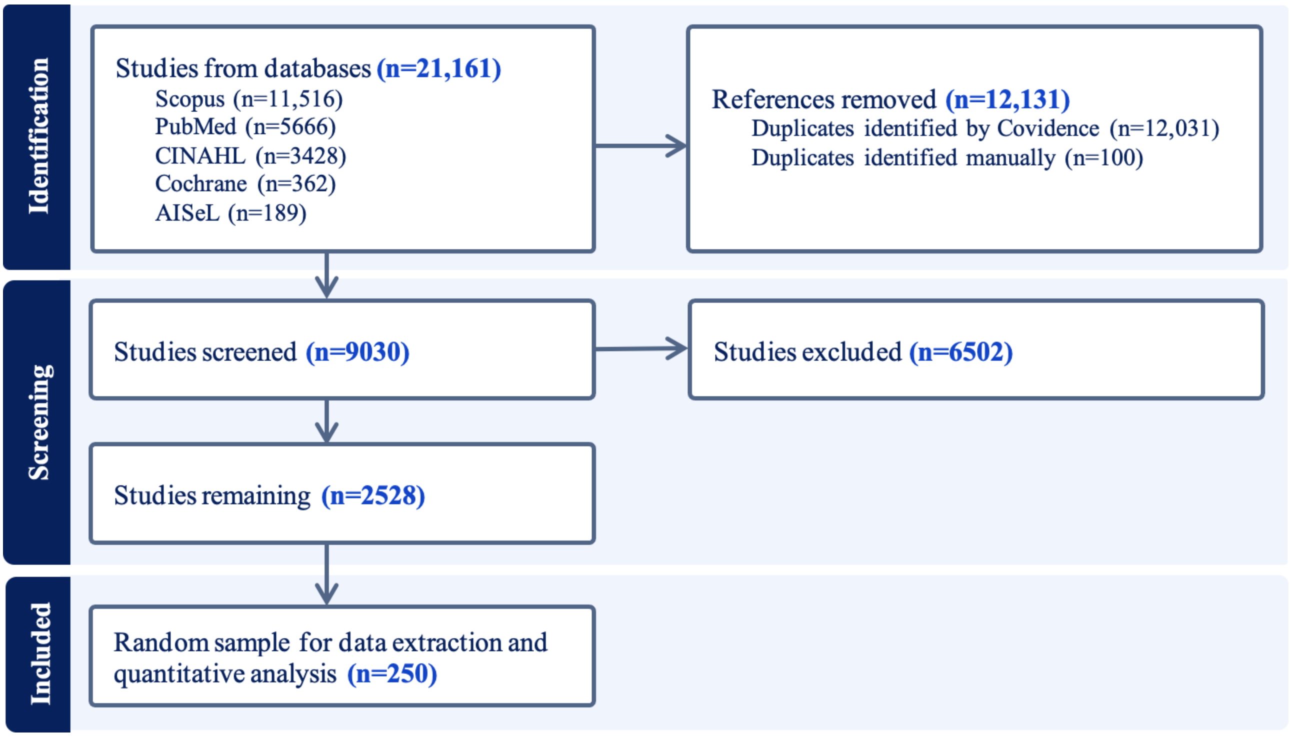
Results

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.

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 (%).

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	

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

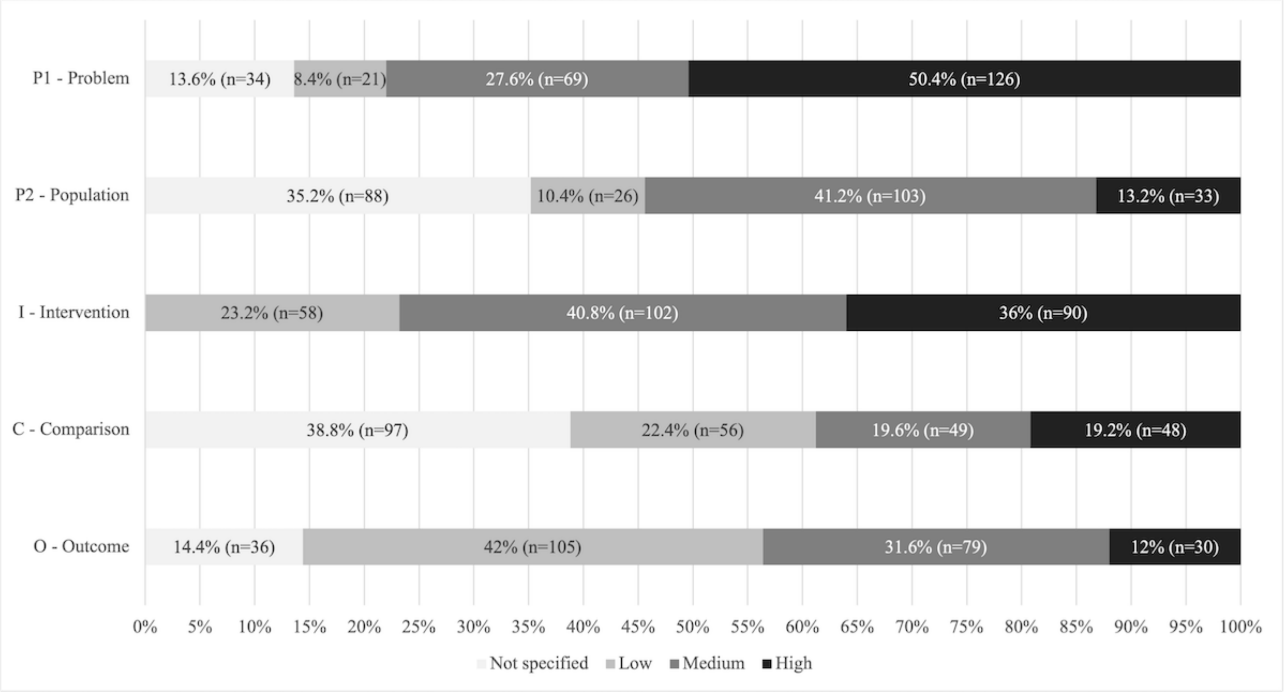
Table 3 and Figure 2 display the level of specification of PICO elements according to the classification scheme (Table S1 in Multimedia Appendix 3) across the sample. A key finding is that no single PICO element was highly specified in more than half of the abstracts.

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.

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;

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.

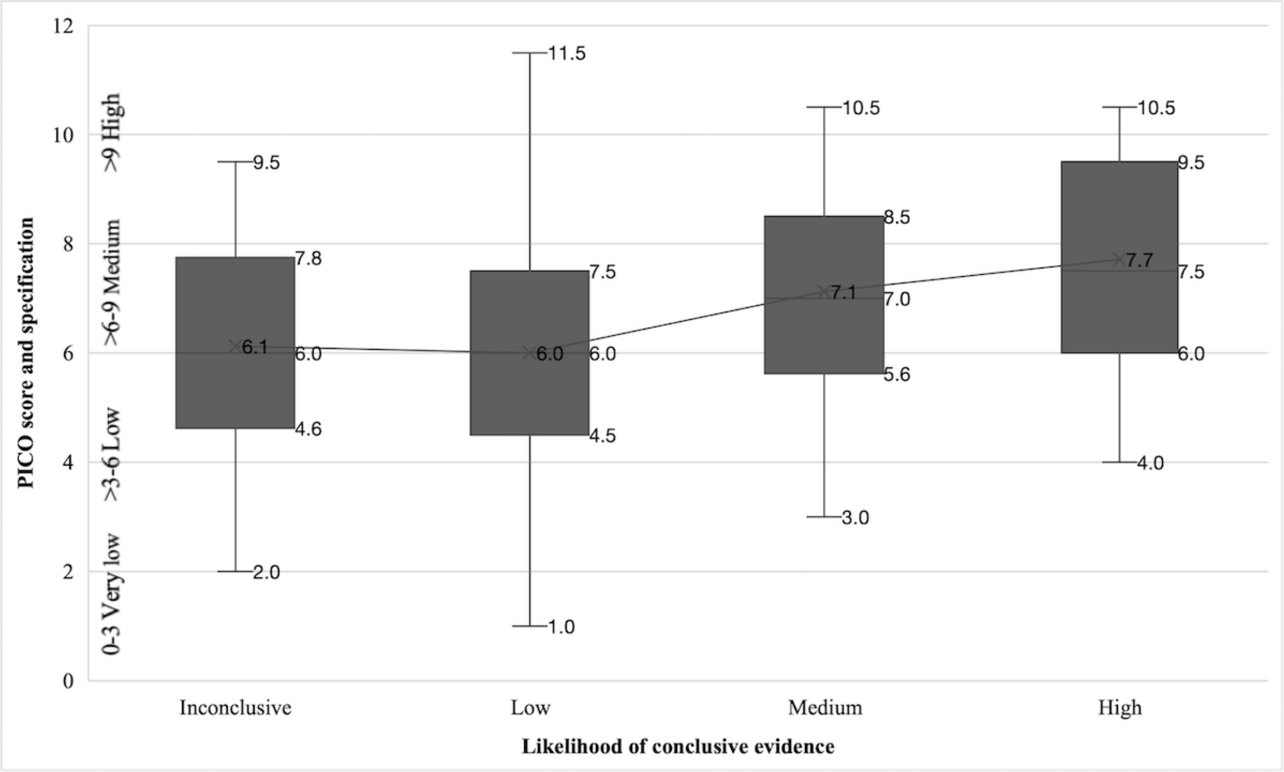
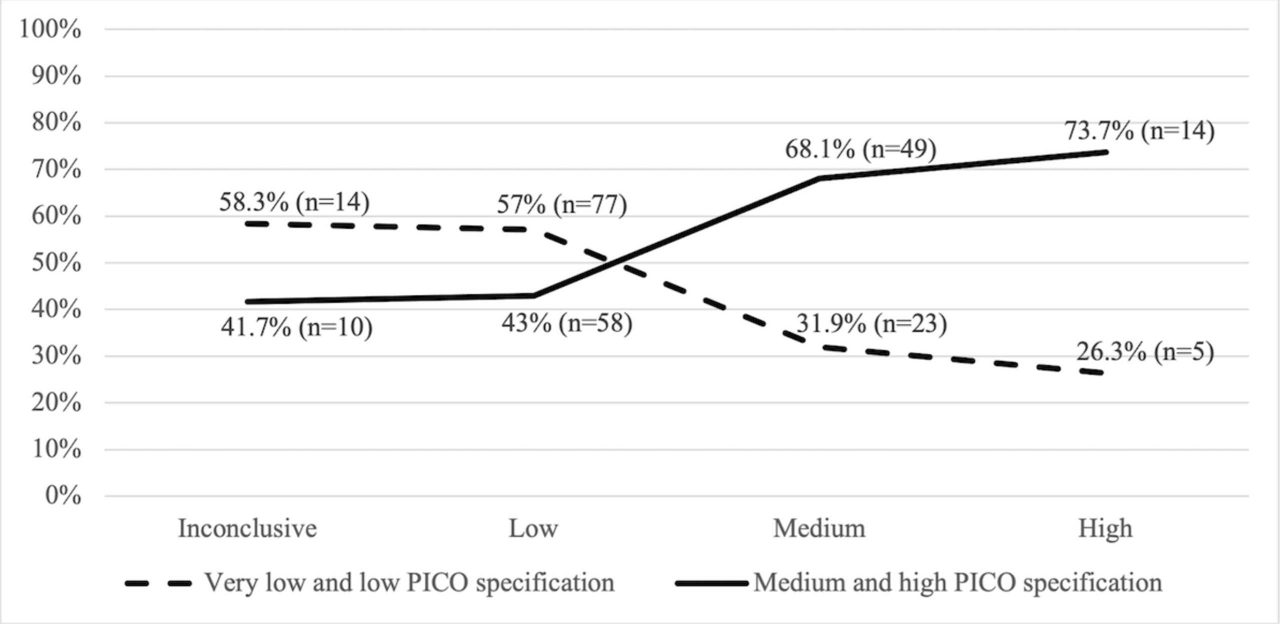


Figure 4. Combined PICO specification shares by likelihood of conclusive evidence classification, % (n) of abstracts in each evidence class. PICO: population, problem, or patient, intervention, comparison, and outcome.



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 (ρ=0.26; *P*<.001) between overall PICO specification (ie, PICO score) and conclusive evidence. Higher specification of the outcome of interest (PICO element O; ρ=0.27; *P*<.001) and of the problem (PICO element P1; ρ=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.

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.

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.

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.

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]

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]

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

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 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

(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.

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

(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).

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						

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

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.

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[a,b].

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 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.

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]

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

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

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], 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].

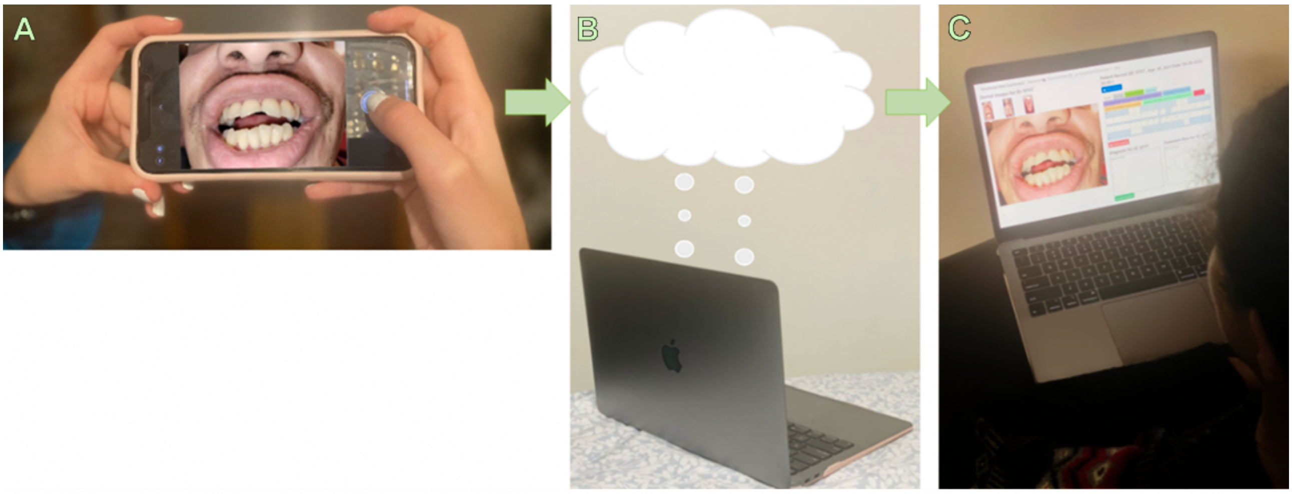
Whether mHealth practice would or would not induce discomfort in dental patients and/or dental professionals remained unclear. Therefore, this study addressed the following research questions: (1) whether dental professionals perceived potential discomfort from their patients in response to their use 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.

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 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

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."

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 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.

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

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]

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

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

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 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.

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.

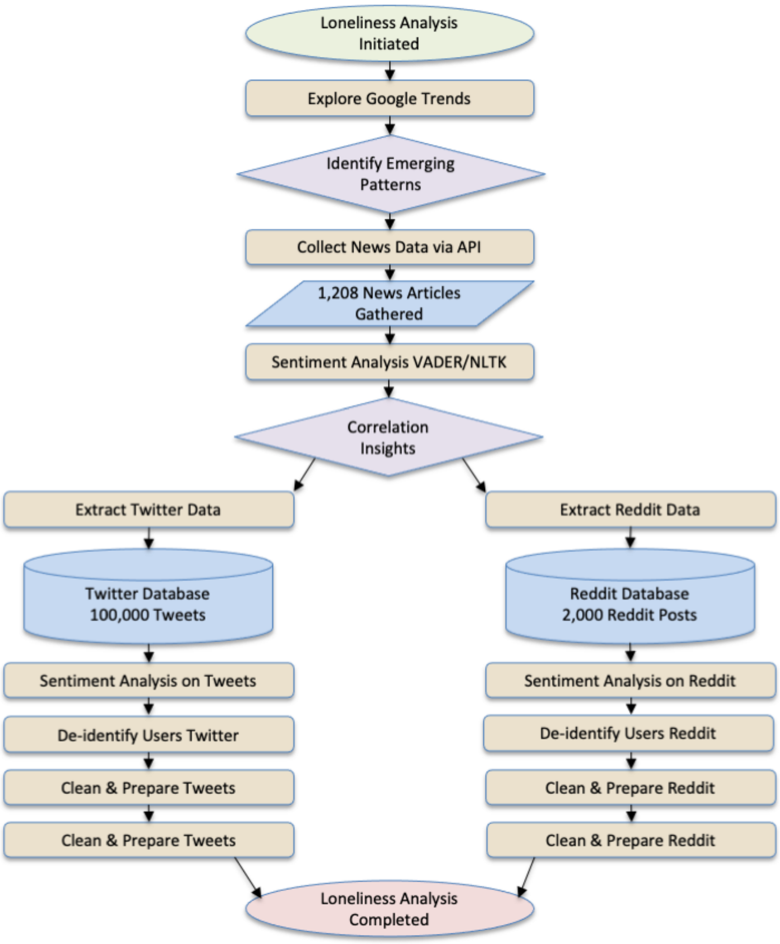
Data Collection

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 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.



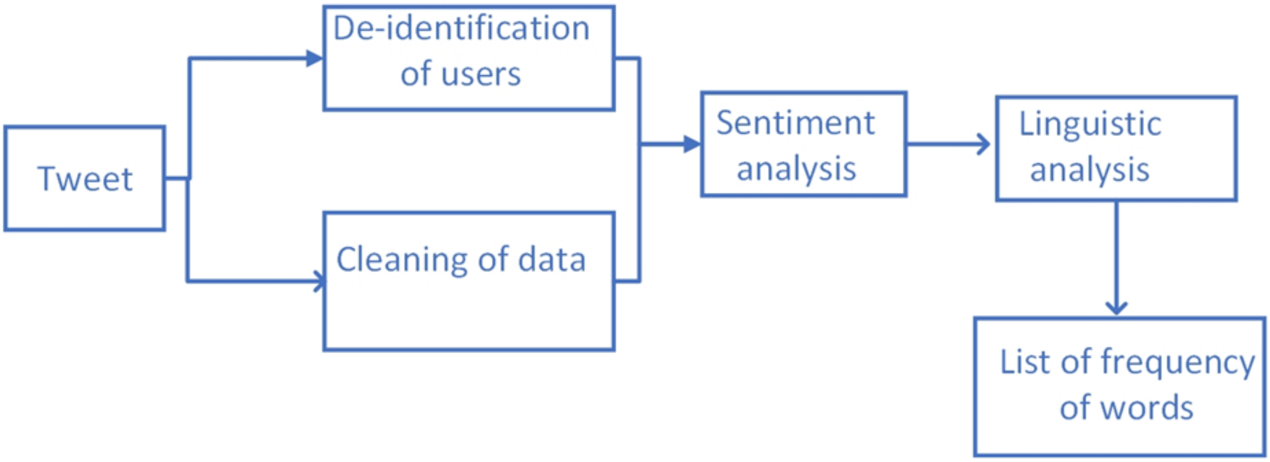
Data Analysis

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 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.



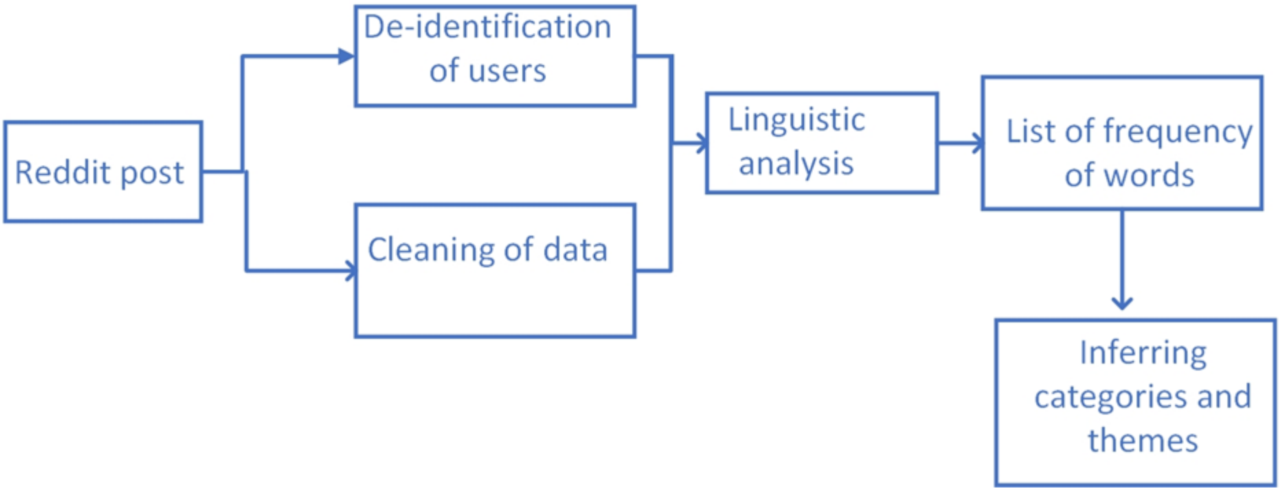
Loneliness in the context of mental health is a negative emotion, which is why the sentiment analysis stage is required—to find out how loneliness is expressed. For an analysis of loneliness in the context of mental health, we filtered out the X posts in which the expression of loneliness was negative. 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.

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 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.

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.

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.

Figure 5. Visualization of highly correlated themes obtained from X posts.

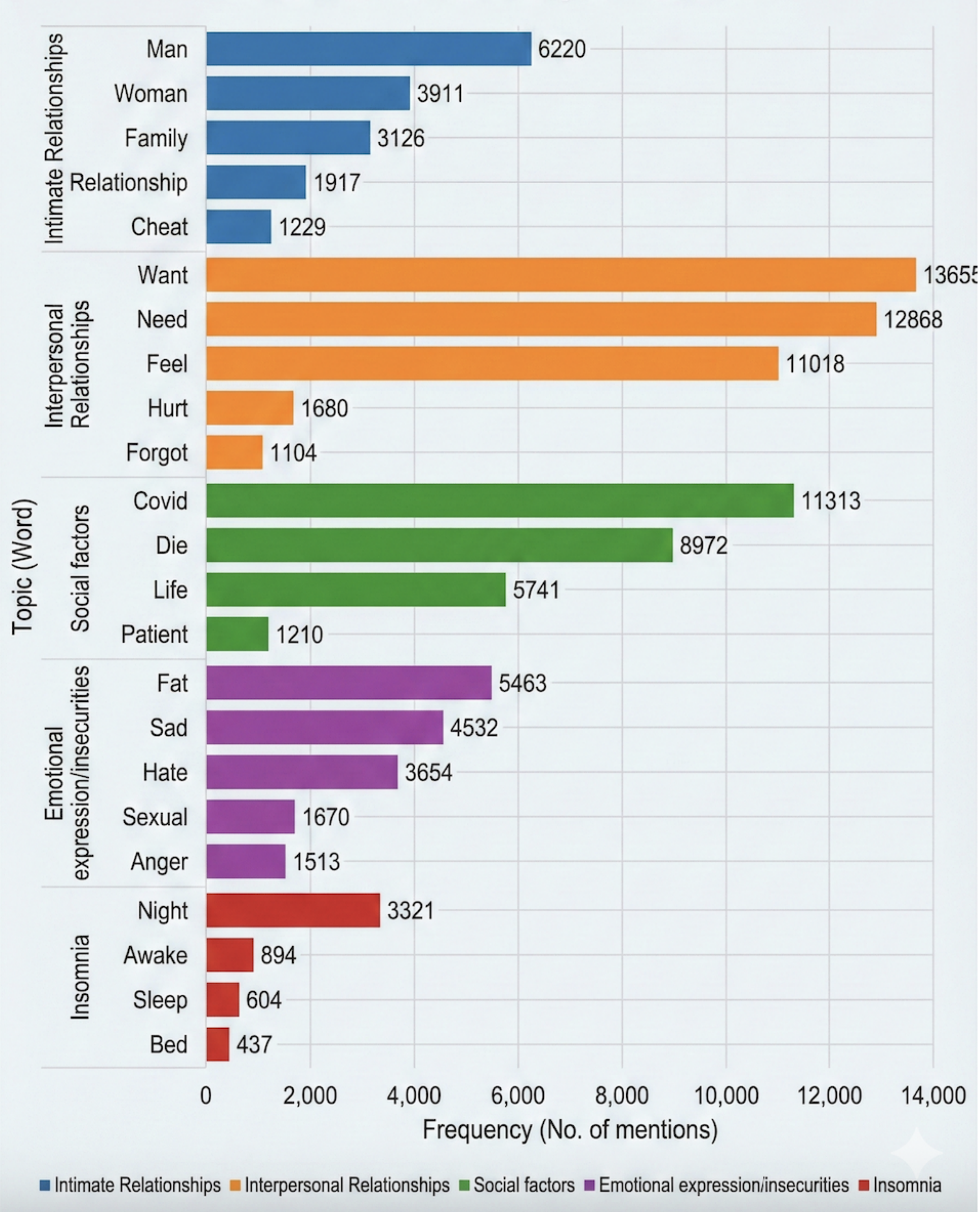


Table 2. Words highly correlated with mentions of loneliness in X posts. Topics are categorized under a broader thematic area.

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)

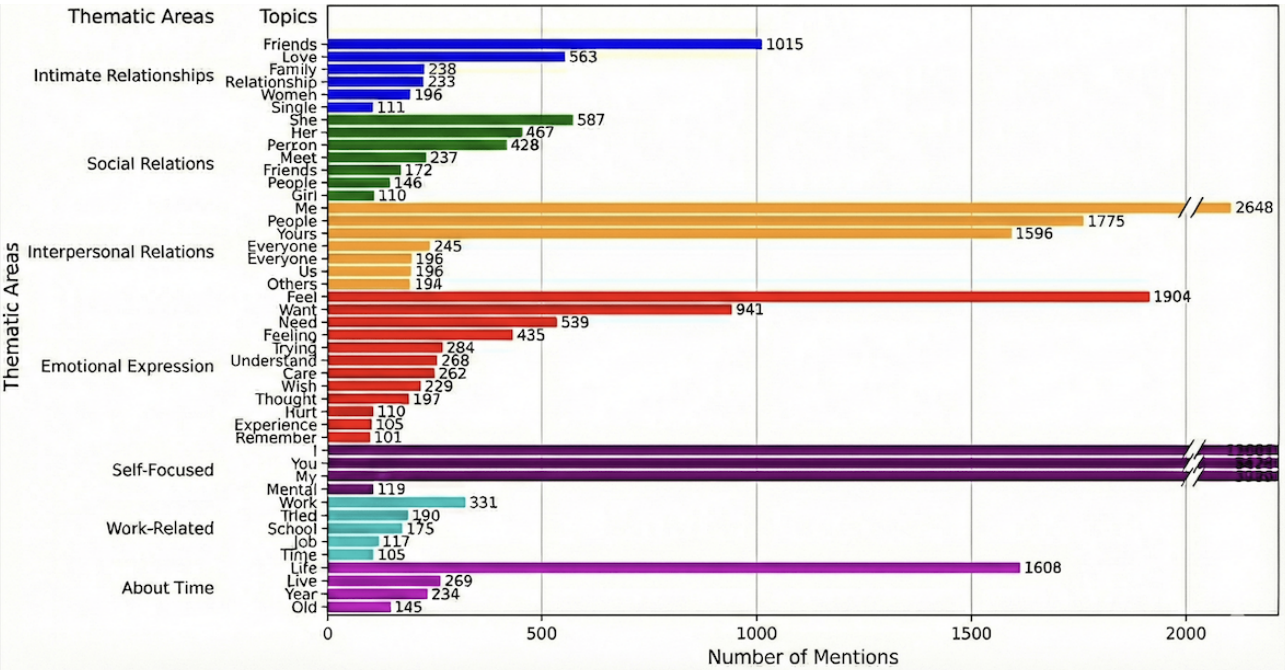
In addition, we want to note that the Reddit posts on loneliness were not specific to the United States. The posts were not divided by country, and the Reddit API does not allow for country-specific downloads. Some methods provide the ability to find the country of the post from the Reddit data, but 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.

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

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 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, 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]

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:

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 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.

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.

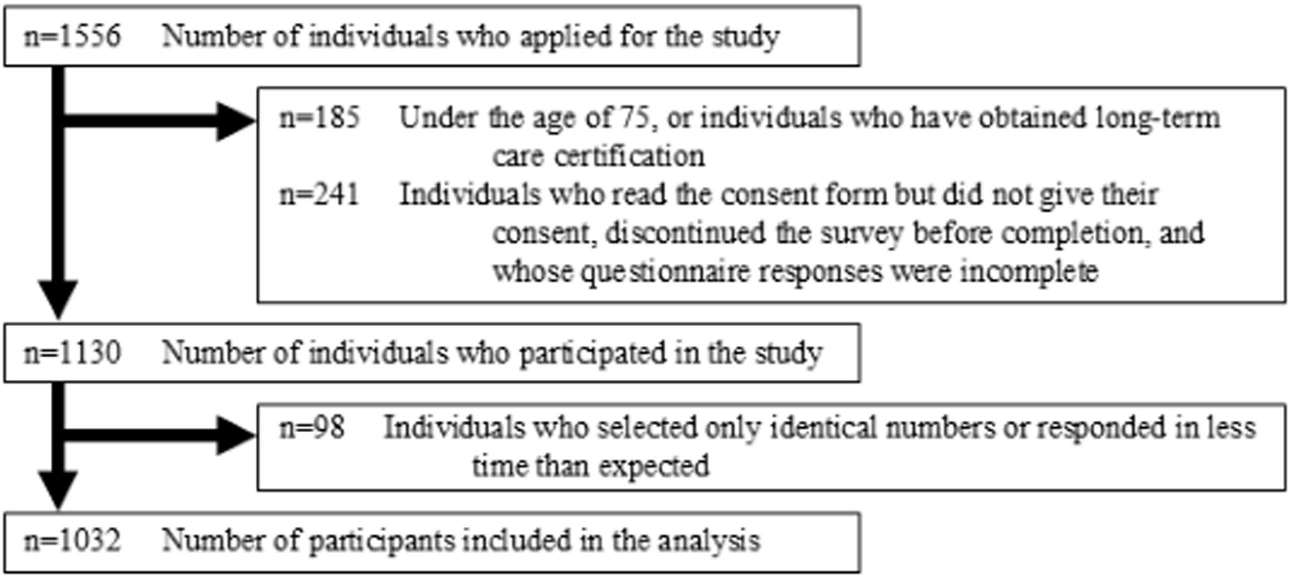


Table . Participant characteristics.

	All participants (N=1032)	Nonfrail (n=714) QMCOO[a] ≤3 points	Frail (n=318) 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)

	All participants (N=1032)	Nonfrail (n=714) QMCOO[a] ≤3 points	Frail (n=318) 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

[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.

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]

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

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

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

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.

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.

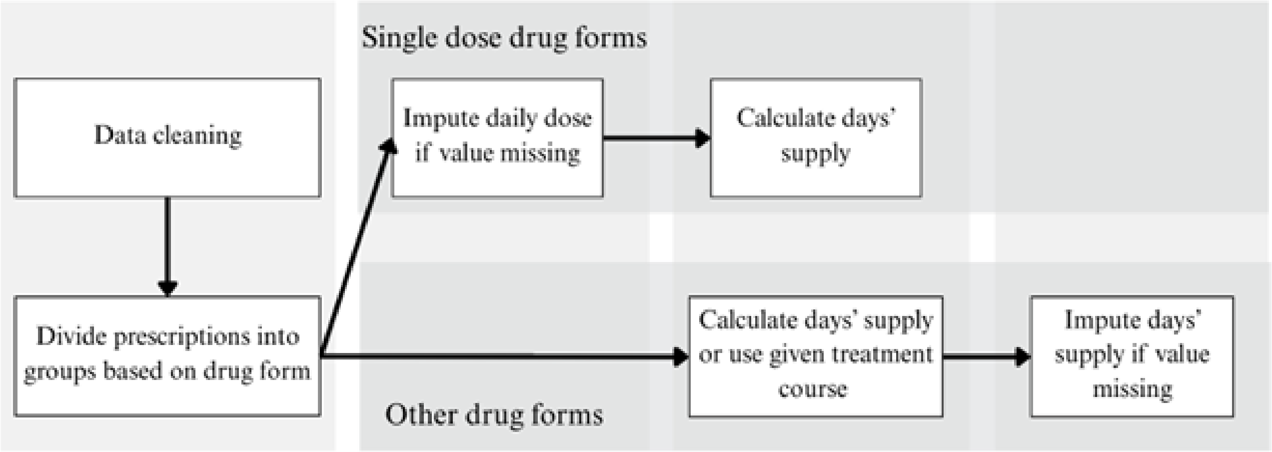
Even though almost all prescriptions after 2016 were provided with a daily dose, the information was inconsistent in terms of the units used. The amount of medication could be given as a quantity (eg, 1 tablet, 2 pills, etc) or the amount of active substance (eg, 10 mg, 20 mg, etc). Neither of these 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.

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

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 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.

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		

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.

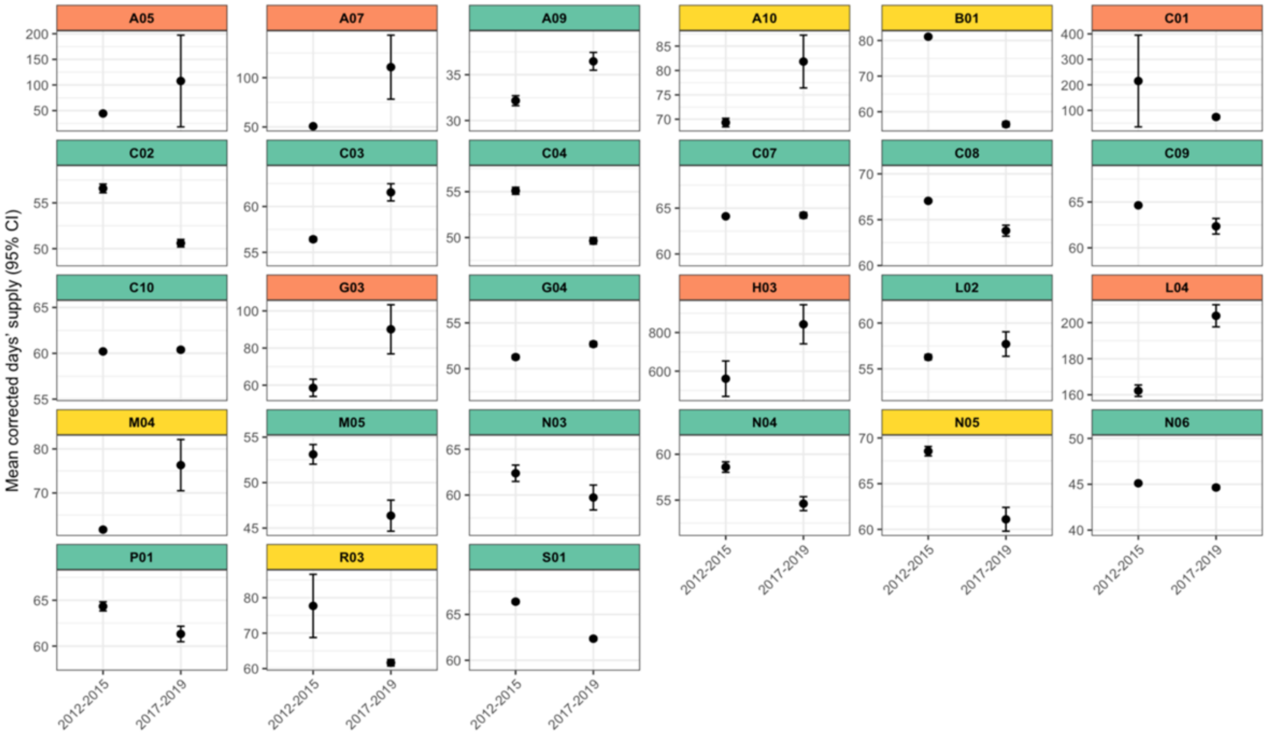
The drug forms with the lowest proportion of correct DS provided were insulin injections (2601/82,867, 3.1%) and intravaginal contraceptives (1692/21,145, 8%) while the highest proportion of DS was provided for inhalation medication (n=78,541/126,588, 62%), oral drops (52,085/98,221, 53%) and tablets, 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.

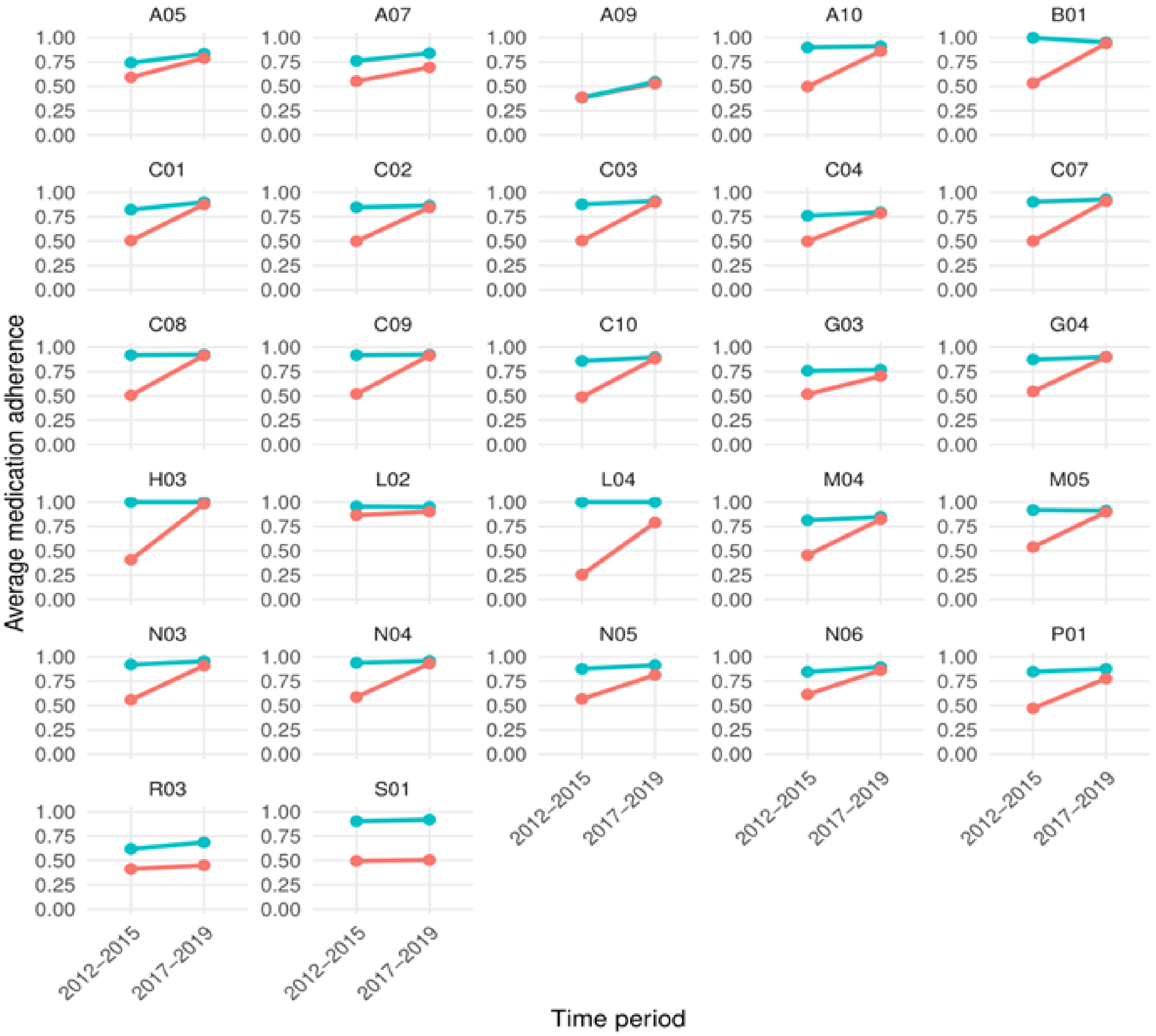
Comparison of the mean DS values between the period where most DS were imputed (2012-2015) and the period with more complete daily dose and DS data (2017-2019) revealed that the observed differences in mean DS were generally small and, in most cases, did not exceed the difference of 7 days (Figure 2). 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).

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.

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 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 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]

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]

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:

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

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.

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, 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.

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.

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.

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.

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 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.

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]

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

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

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].

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)

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.

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).

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.

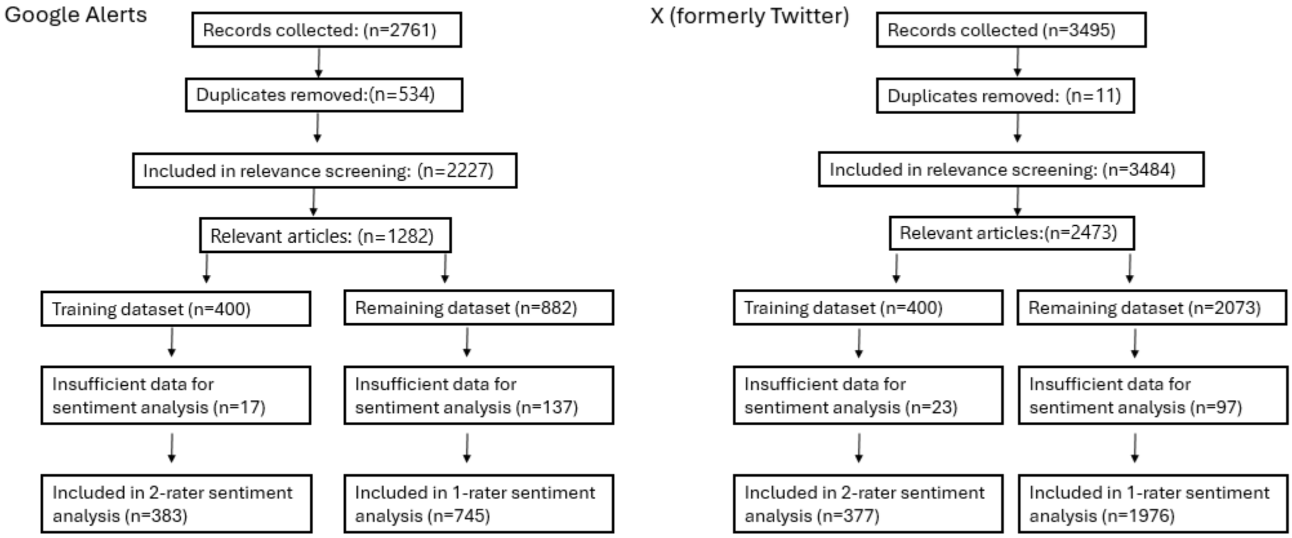
Description of the Dataset

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

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.

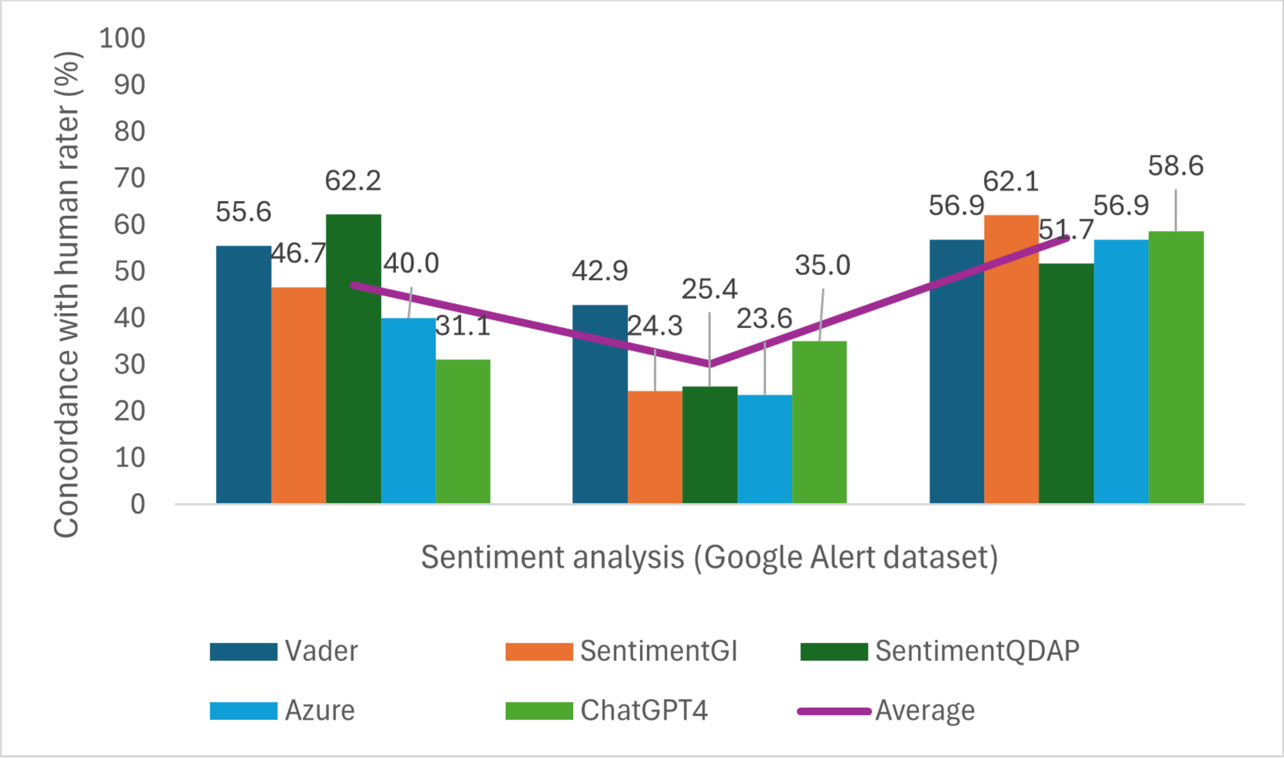
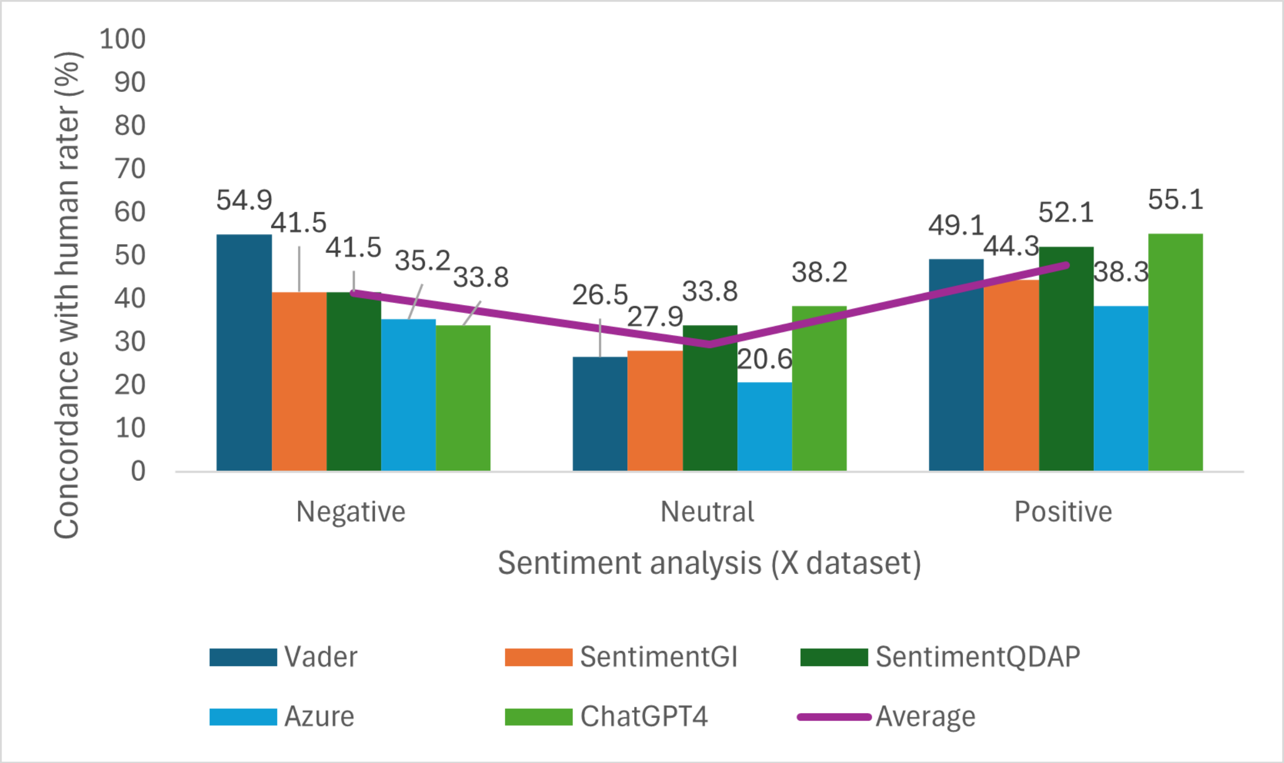


Figure 3. Comparison of human-rated sentiment score for a subset of tweets (n=377) with 5 artificial intelligence tools.



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

"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.

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.

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").

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.

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

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.

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

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.

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

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.

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

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 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 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]

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]

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]

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

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

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:

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 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.

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)

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.

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 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;

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)

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.

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 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]

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

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

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 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 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, 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.

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.

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.

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]

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

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

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 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.

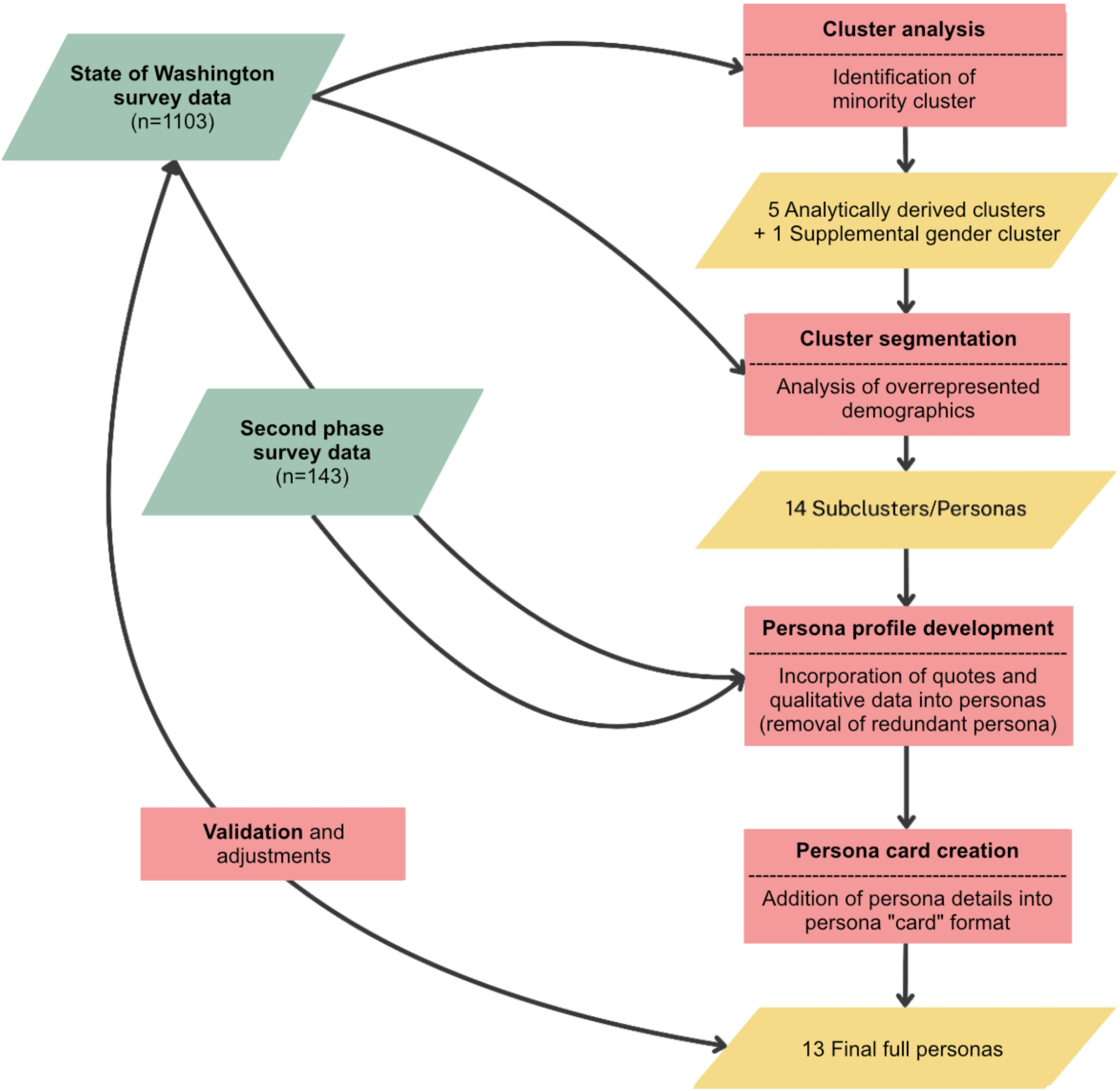
The second-phase survey targeted a smaller sample but aimed to gather more specific information about barriers to technology use in the public health context. The survey focused on digital literacy, internet safety perceptions, digital privacy and security, confidence in public health initiatives, and 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.

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.

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 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 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 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.

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)

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

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.

Figure 2. Persona 5: Desmond.

Desmond

Technically savvy and embraces the convenience technology offers

Occupation: Banking
Age: 37
Race/Ethnicity: Black
English Only Home: Yes
Region: Western WA
Education: 4-year
Has Minor Child: No
Gender: Male

Public Health & Tech Characteristics

- ↑ High Level of Support for PH Policies
- ✓ Owns a Smartphone
- High Tech Readiness
- Uses Online Health Tools

Drivers

- Believes technology makes everything convenient but is wary of sharing data online
- Has witnessed malicious use of stolen information through his work
- Values personal safety and uses tools that will support his own and his family's safety
- Aware of the danger of algorithms perpetuating bias

Quotes

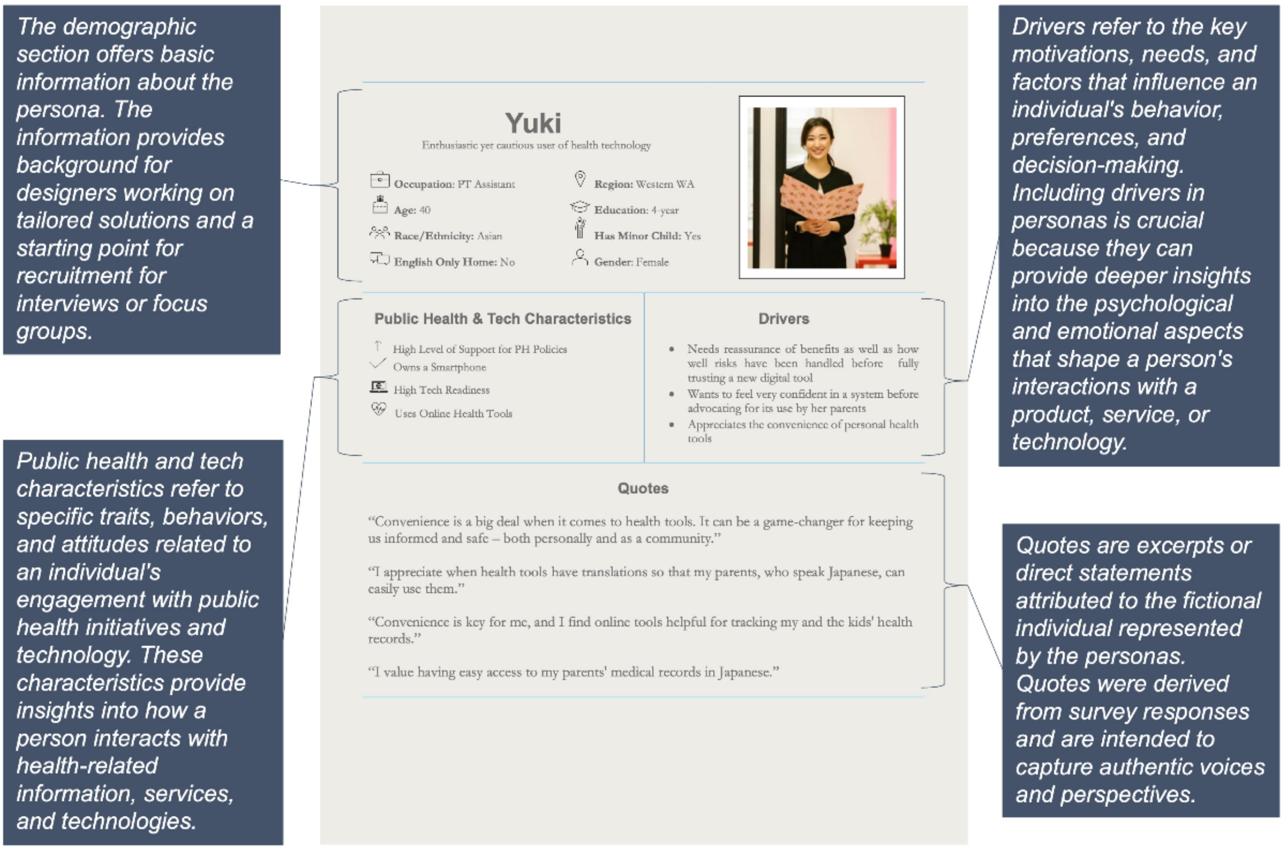
"Data and technology can be so helpful and really help make things faster and more convenient in everyday life. The problem is that there are aways people with malicious intent out there who can use any information they get their hands on to cause harm."

"I only have one dedicated credit card that I use online and I track it regularly. I am hesitant to create new accounts and share health or other personal information online. I weigh the potential risks with the benefits to make a decision about any new app or account."

"I am often out and about and since my wife has an autoimmune disease, I really value the security of knowing that there is some level of safeguard around us. To us, downloading and using online public health tools are usually worth it."

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.

Figure 3. Sample persona with section descriptions.



The demographic section offers basic information about the persona. The information provides background for designers working on tailored solutions and a starting point for recruitment for interviews or focus groups.

Yuki

Enthusiastic yet cautious user of health technology

Occupation: PT Assistant
Age: 40
Race/Ethnicity: Asian
English Only Home: No
Region: Western WA
Education: 4-year
Has Minor Child: Yes
Gender: Female

Public Health & Tech Characteristics

↑ High Level of Support for PH Policies
✓ Owns a Smartphone
High Tech Readiness
Uses Online Health Tools

Drivers

- Needs reassurance of benefits as well as how well risks have been handled before fully trusting a new digital tool
- Wants to feel very confident in a system before advocating for its use by her parents
- Appreciates the convenience of personal health tools

Quotes

"Convenience is a big deal when it comes to health tools. It can be a game-changer for keeping us informed and safe – both personally and as a community."

"I appreciate when health tools have translations so that my parents, who speak Japanese, can easily use them."

"Convenience is key for me, and I find online tools helpful for tracking my and the kids' health records."

"I value having easy access to my parents' medical records in Japanese."

Drivers refer to the key motivations, needs, and factors that influence an individual's behavior, preferences, and decision-making. Including drivers in personas is crucial because they can provide deeper insights into the psychological and emotional aspects that shape a person's interactions with a product, service, or technology.

Public health and tech characteristics refer to specific traits, behaviors, and attitudes related to an individual's engagement with public health initiatives and technology. These characteristics provide insights into how a person interacts with health-related information, services, and technologies.

Quotes are excerpts or direct statements attributed to the fictional individual represented by the personas. Quotes were derived from survey responses and are intended to capture authentic voices and perspectives.

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 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 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

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]

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

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

[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 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.

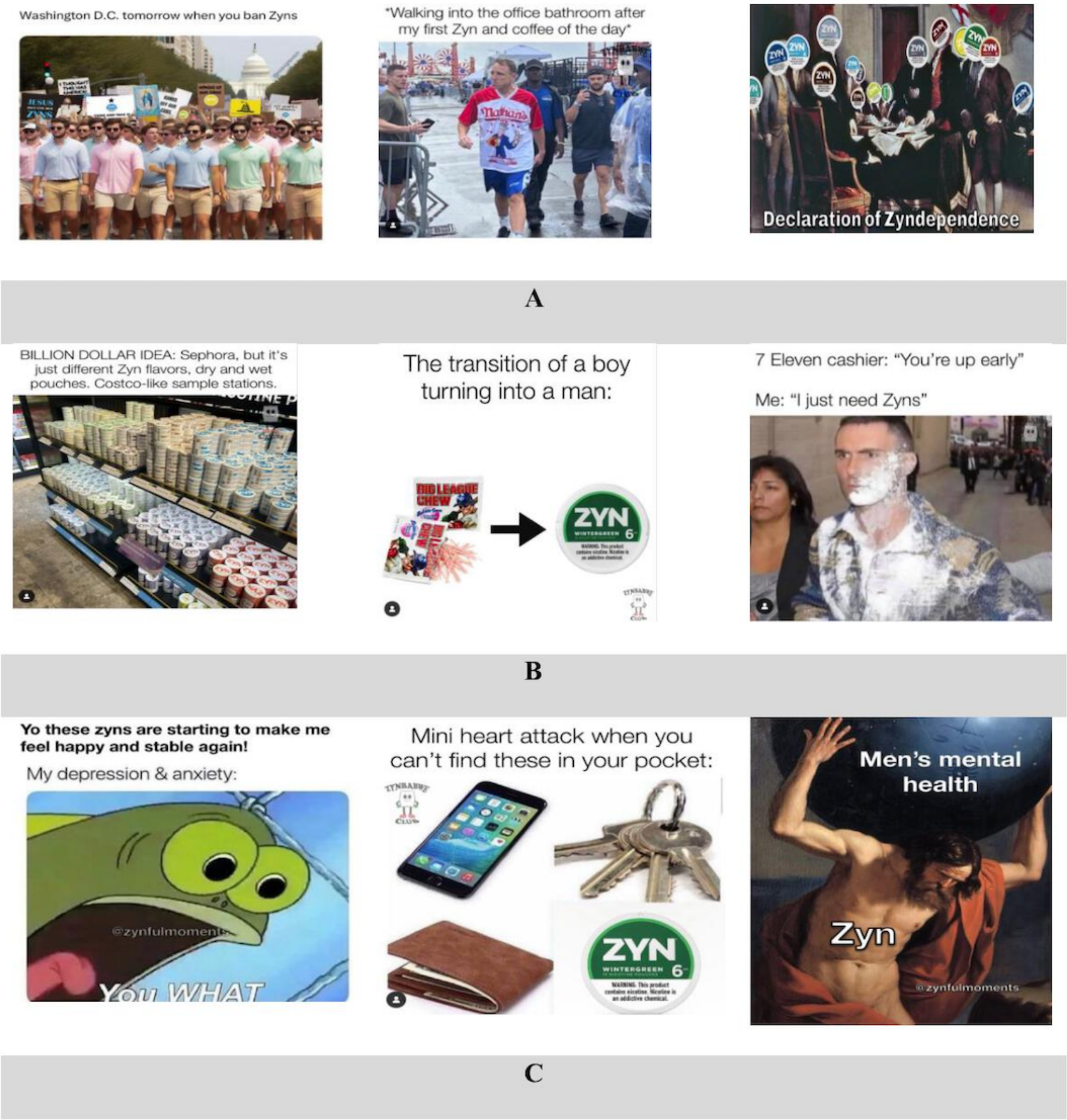
Engagement Metrics for the Top 3 Prominent Themes

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.

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 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.

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].

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 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]

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]

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

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

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 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 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.

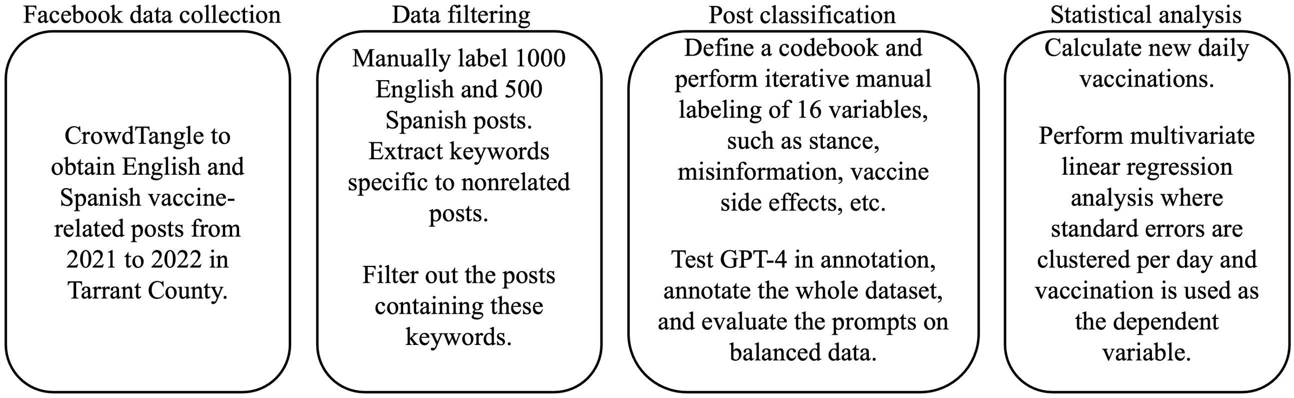
Language was also included as a control variable coded as English or Spanish. This allowed the analysis to test whether post language was associated with differential vaccination uptake while recognizing that language does not map perfectly onto ethnicity.

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

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

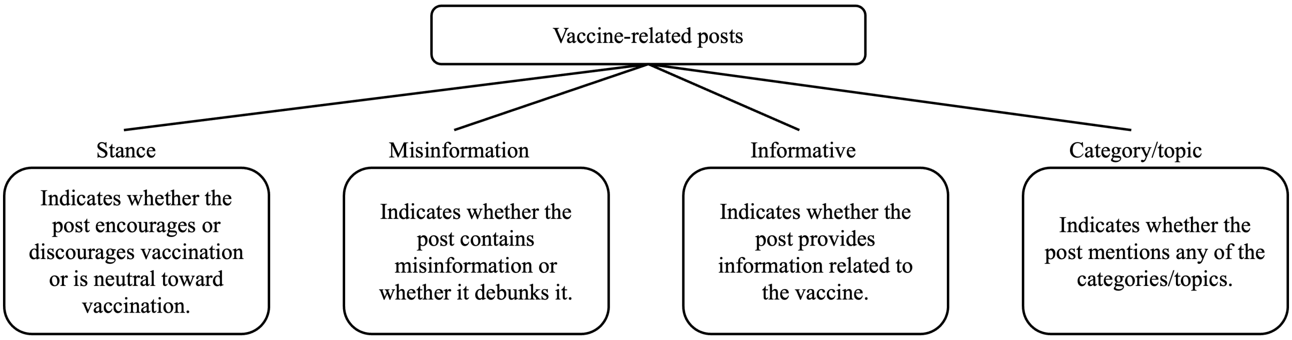
Although querying based on vaccine-related keywords yielded many Facebook posts, an essential step before further data processing involved ensuring that the posts obtained were related to COVID-19 vaccines. Initially, a random sample of 1000 English posts and 500 Spanish posts was extracted from the 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 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 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.

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, 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.

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.

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.

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 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.

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]

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]

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]

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]

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

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

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].

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].

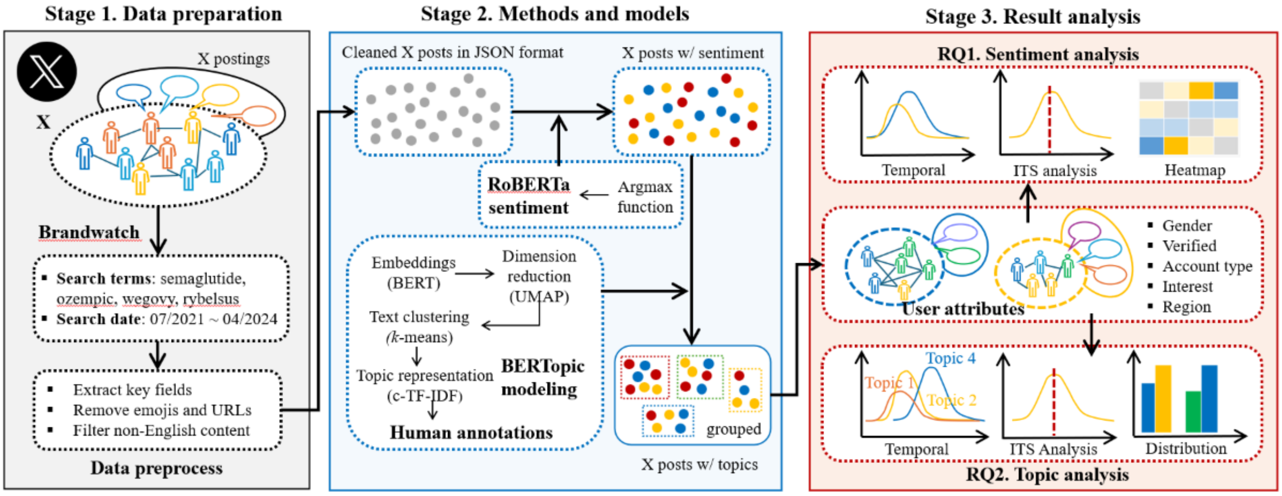
Topic modeling techniques, such as latent Dirichlet allocation [41] and BERTopic [42], seek to discover the key themes present in a corpus of documents [43]. Topic models can summarize large datasets by capturing the topics (ie, the major discourse) that appear most commonly in the text. Numerous works 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].

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 $Time_t Event_t$ 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 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.

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)

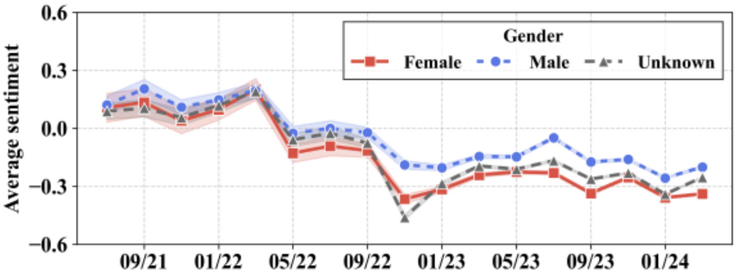
Addressing RQ1, we find that the overall sentiment toward semaglutide during the study period (July 2021 to April 2024) is slightly negative, with a mean sentiment score of –0.24 (SD 0.669) across all posts and –0.28 (SD 0.605) across all users. Within all user subpopulations, we observe a decline in sentiment over time, but the trend varies across different subpopulations, as discussed in the following subsections.

Overall Sentiment Declines

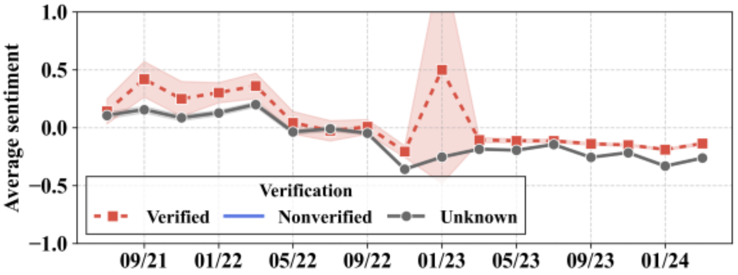
Our longitudinal sentiment analysis results are displayed in Figure 2. To illustrate uncertainty, we include 95% CIs for key categorical attributes as shaded bands. The temporal progression of sentiment can be divided into four phases: (1) initial positive sentiment across categories (2021 to mid-2022), (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.

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.

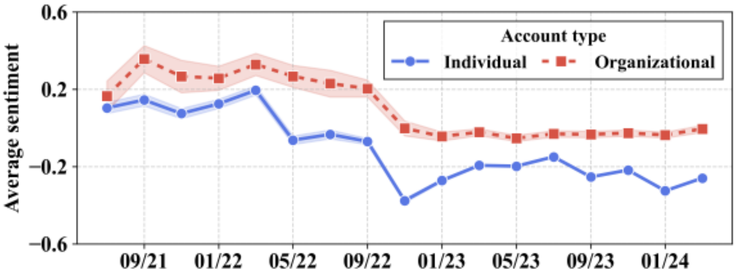
(A) Bimonthly sentiment analysis by gender



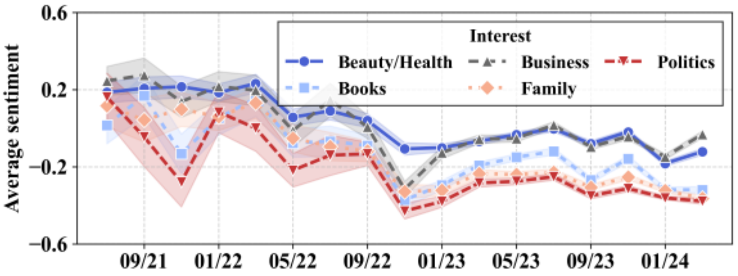
(B) Bimonthly sentiment analysis by verified status



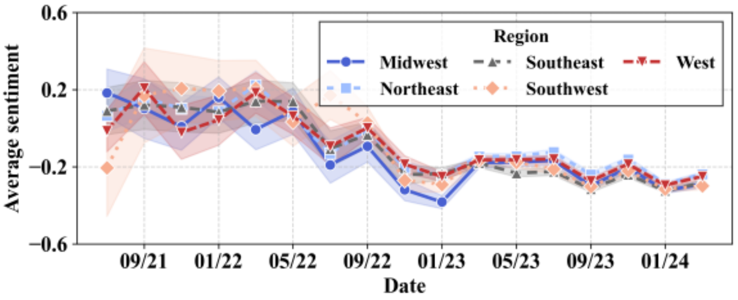
(C) Bimonthly sentiment analysis by account type



(D) Bimonthly sentiment analysis by interest



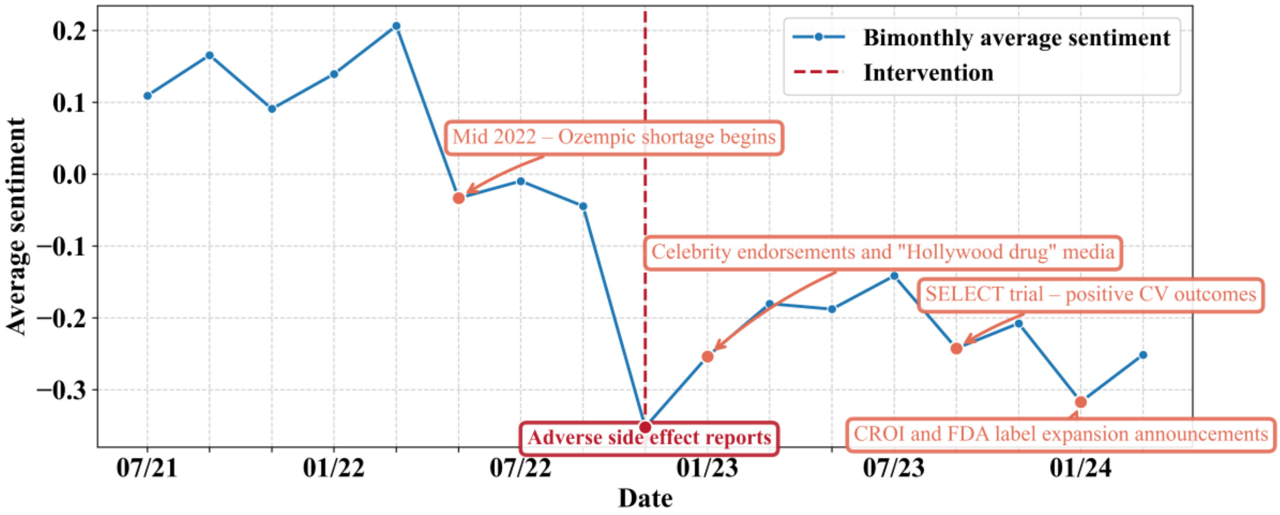
(E) Bimonthly sentiment analysis by US region



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 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.

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).

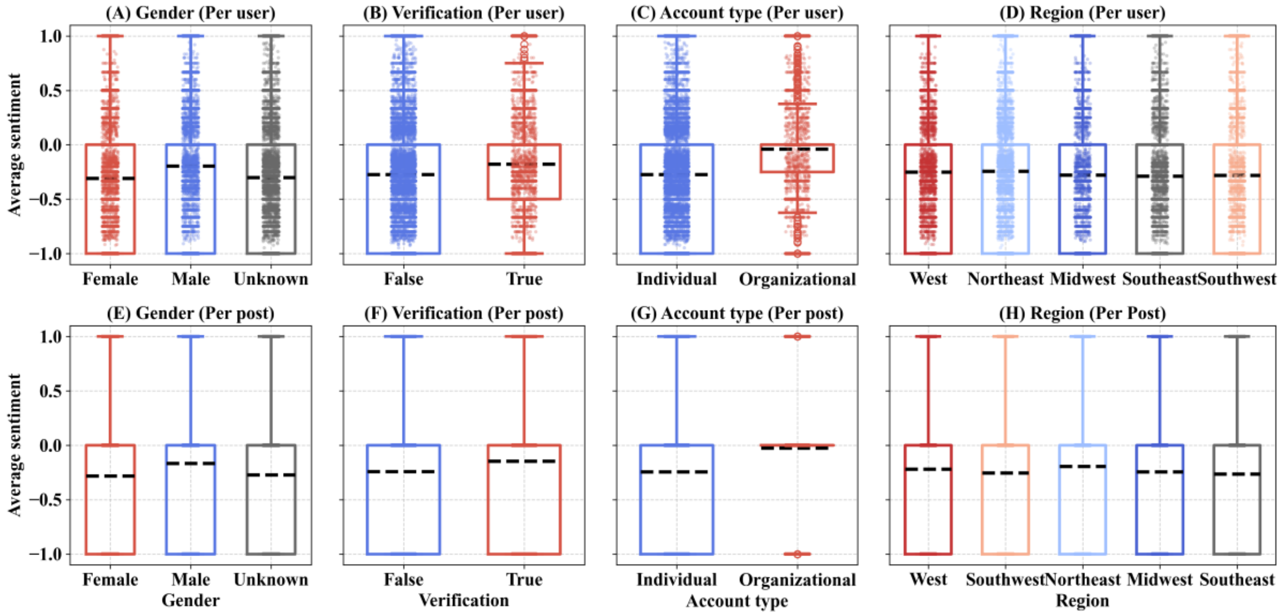


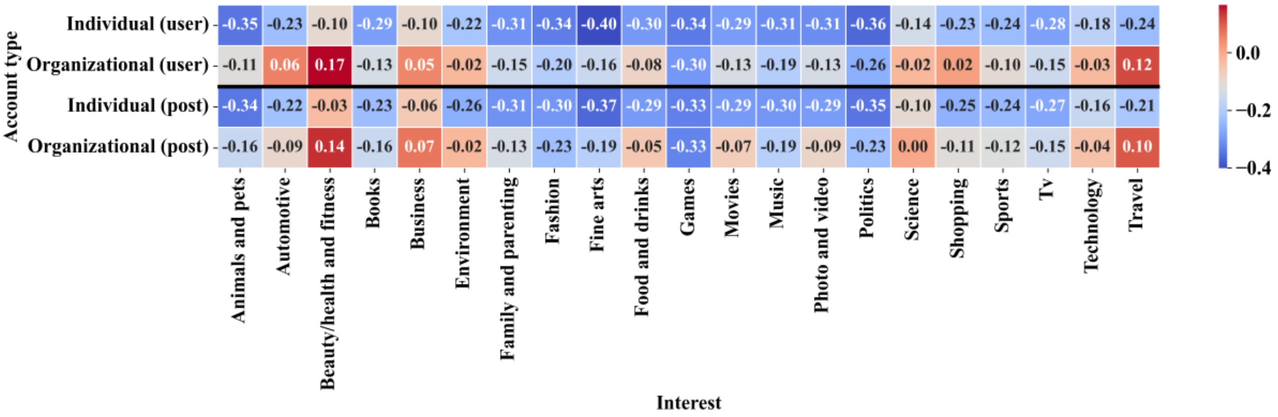
Table 3. Estimated differences in sentiment across key subgroups, with 95% CIs.

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.



	Animals and pets	Automotive	Beauty/health and fitness	Books	Business	Environment	Family and parenting	Fashion	Fine arts	Food and drinks	Games	Movies	Music	Photo and video	Politics	Science	Shopping	Sports	Tv	Technology	Travel
Individual (user)	-0.35	-0.23	-0.10	-0.29	-0.10	-0.22	-0.31	-0.34	-0.40	-0.30	-0.34	-0.29	-0.31	-0.31	-0.36	-0.14	-0.23	-0.24	-0.28	-0.18	-0.24
Organizational (user)	-0.11	0.06	0.17	-0.13	0.05	-0.02	-0.15	-0.20	-0.16	-0.08	-0.30	-0.13	-0.19	-0.13	-0.26	-0.02	0.02	-0.10	-0.15	-0.03	0.12
Individual (post)	-0.34	-0.22	-0.03	-0.23	-0.06	-0.26	-0.31	-0.30	-0.37	-0.29	-0.33	-0.29	-0.30	-0.29	-0.35	-0.10	-0.25	-0.24	-0.27	-0.16	-0.21
Organizational (post)	-0.16	-0.09	0.14	-0.16	0.07	-0.02	-0.13	-0.23	-0.19	-0.05	-0.33	-0.07	-0.19	-0.09	-0.23	0.00	-0.11	-0.12	-0.15	-0.04	0.10

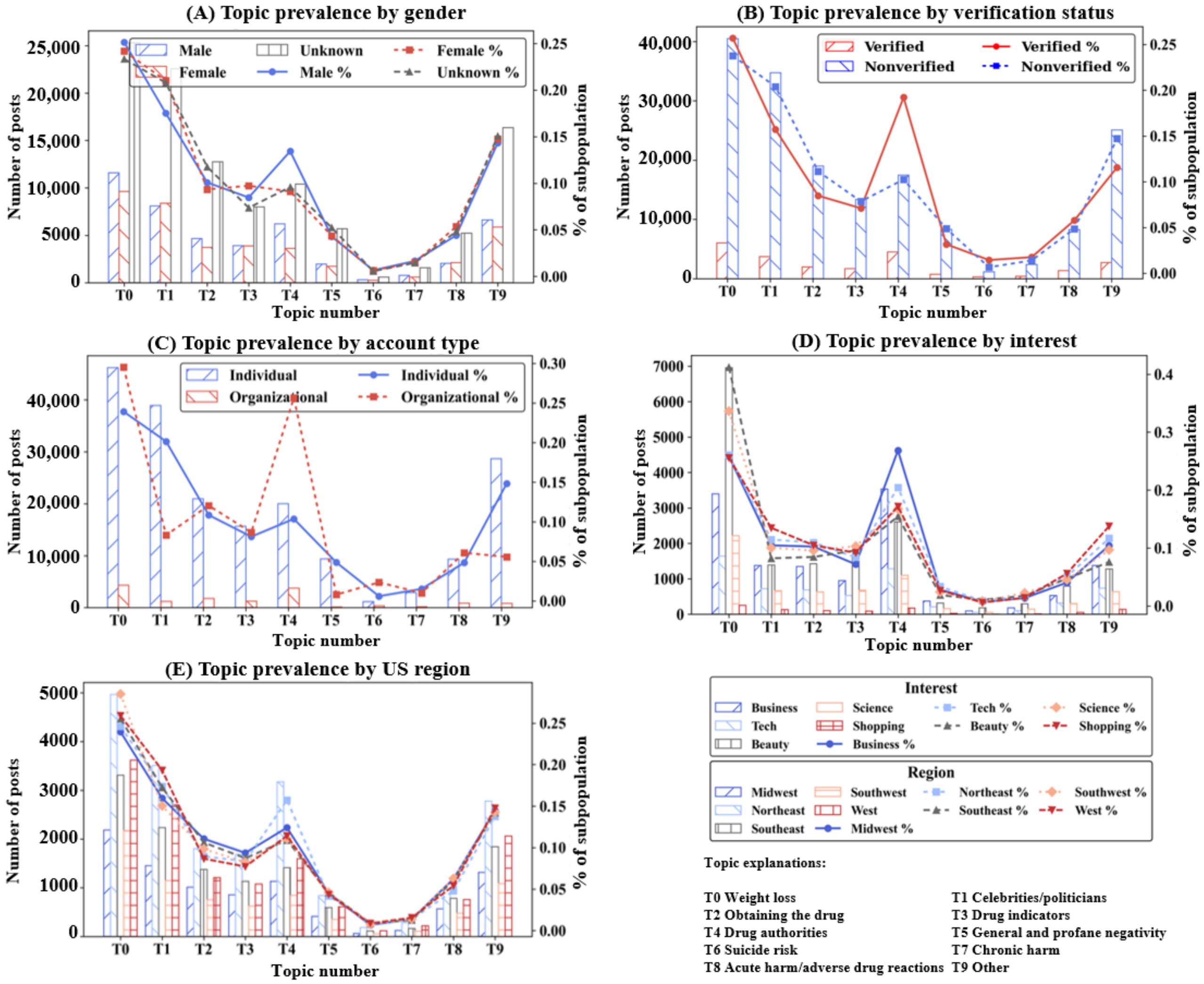
RQ2 (Topic Results Across User Subpopulations)

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.

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.

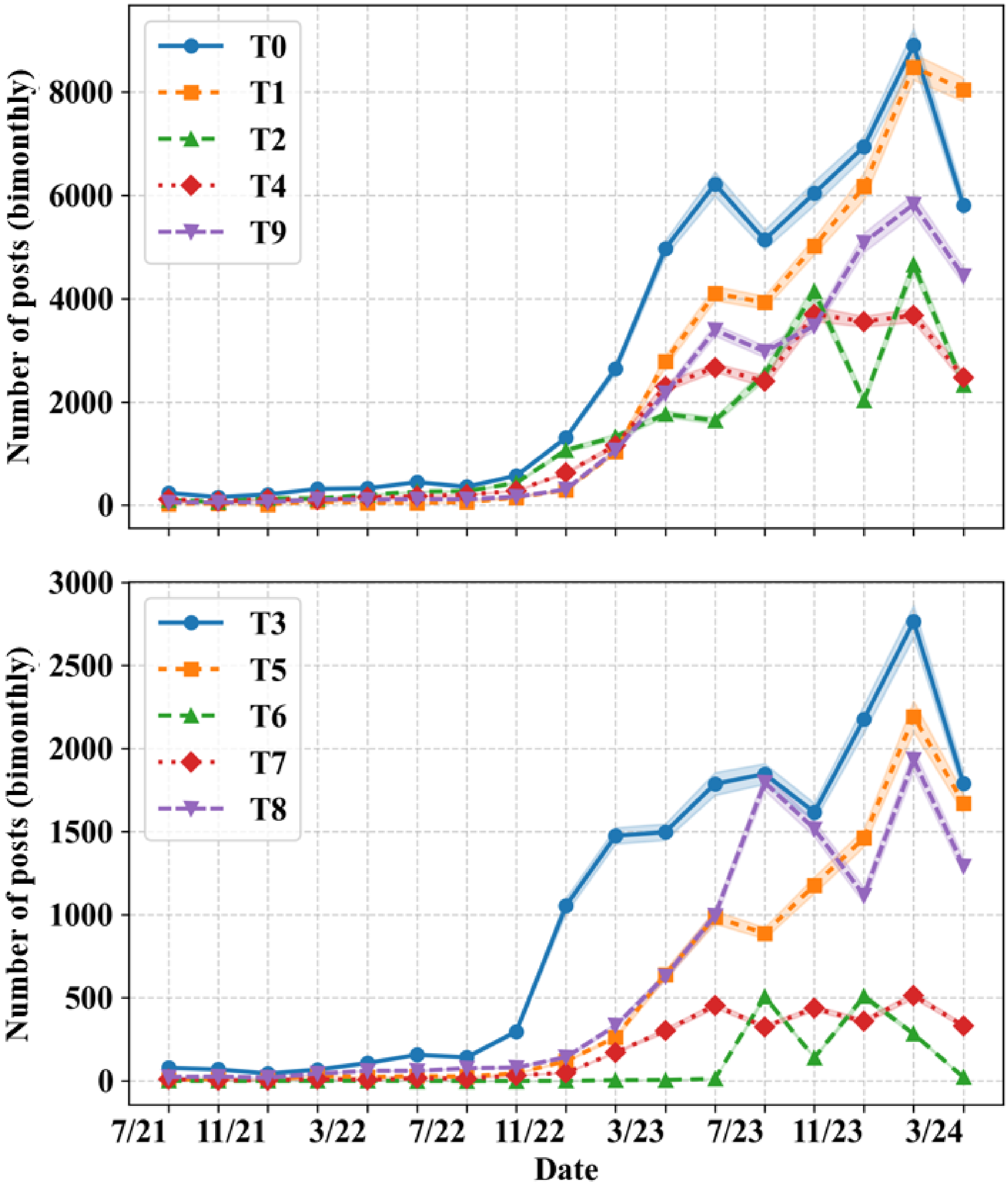
Examining the most prevalent topics among individual accounts and organizational accounts, the most striking difference is the very low number of profane posts within the organizational account subpopulation. T5 comprises just 0.8% (117/14,711) of organizational posts, compared to 4.9% (9371/193,138) 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.

Topic Prevalence Over Time

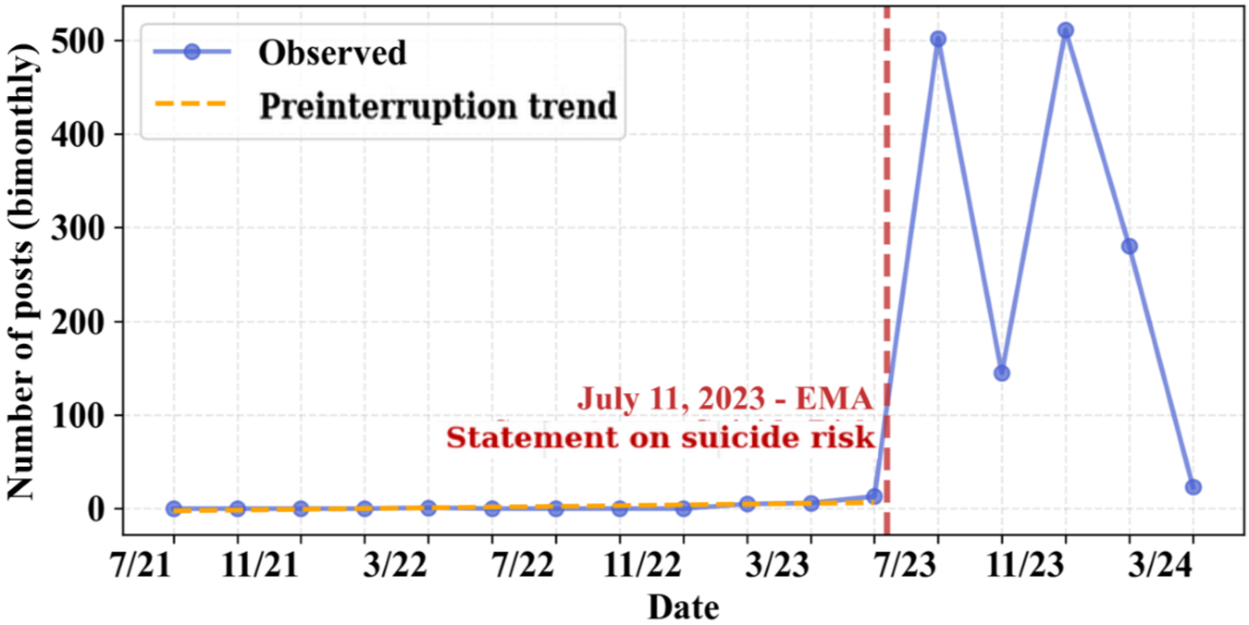
Figure 7 shows the number of posts pertaining to each umbrella topic posted during each bimonthly period from July 2021 to April 2024. To assess the evolution of topic prevalence over time, we first present an exploratory analysis of external events that may have influenced the topic trends. All topics 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 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].

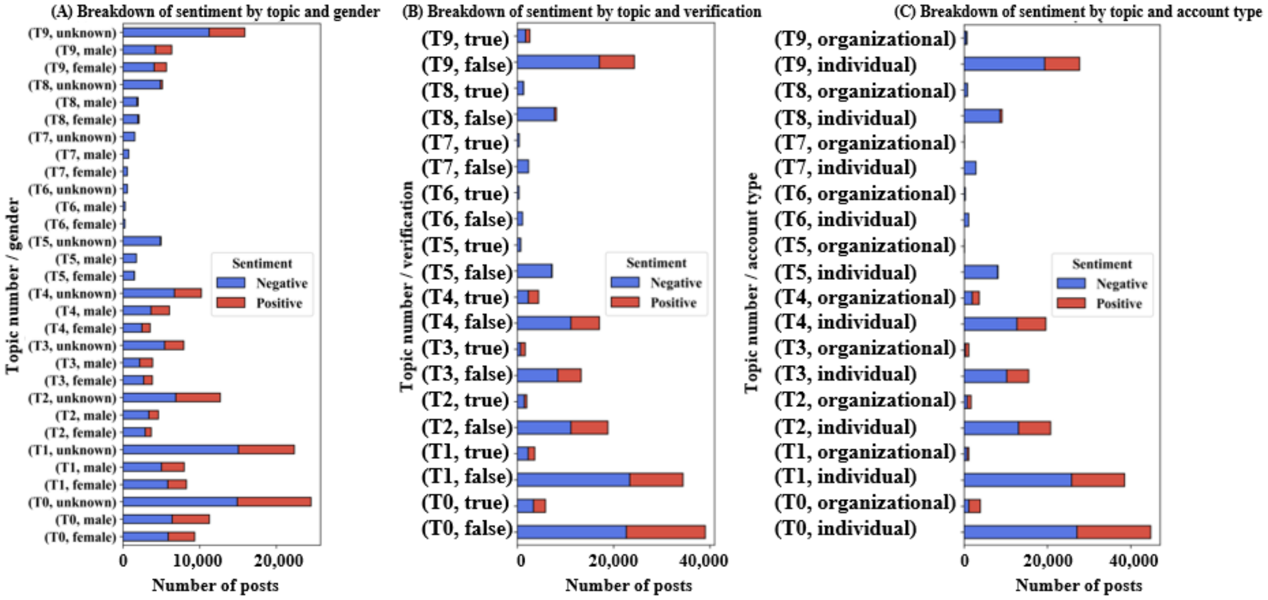
Figure 8. Interrupted time series analysis of T6 (suicide risk). The intervention window is July 2023–September 2023. EMA: European Medicines Agency.



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 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 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]

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]

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]

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]

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:

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 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.

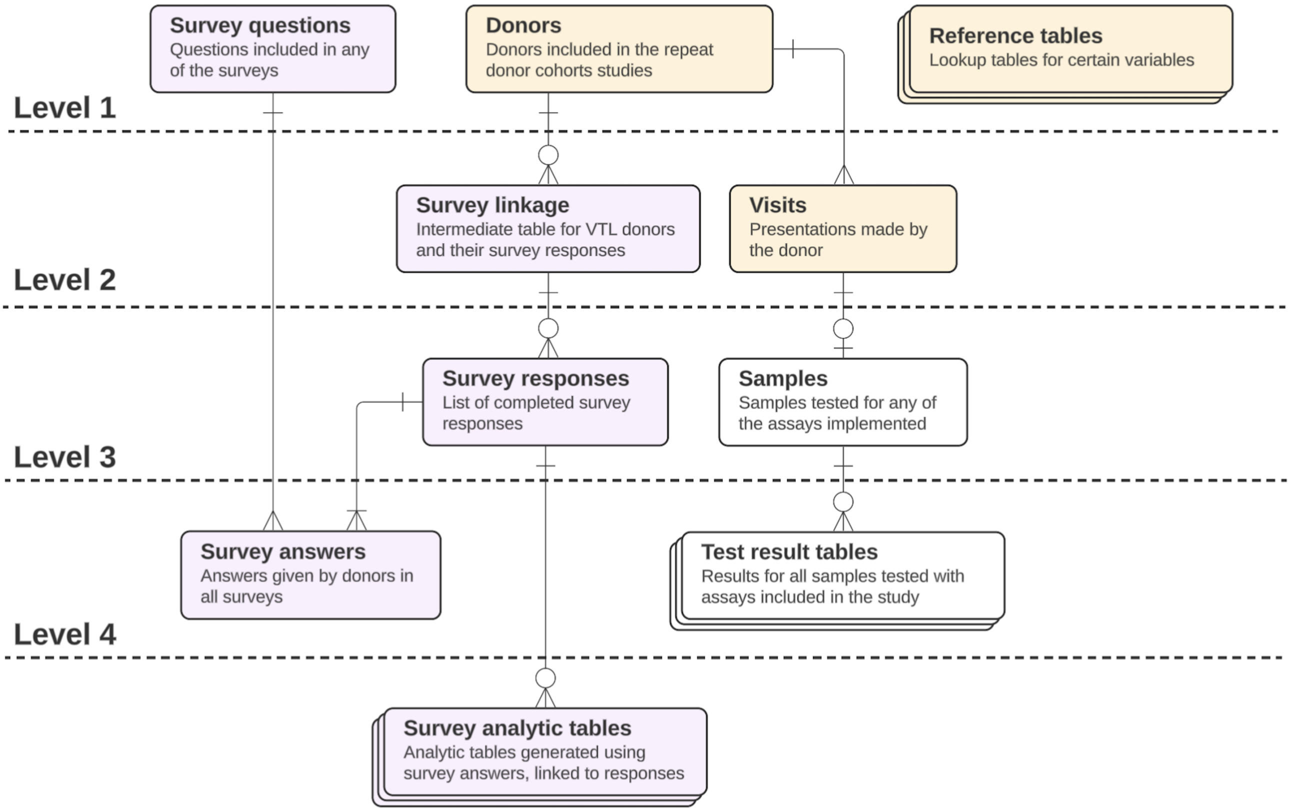
The modules are platform-agnostic and can be run on either a cloud platform or locally with little to no input, depending on how much of the process requires manual review. They could also be run on a set schedule as a *cron job*, a system for scheduling specific tasks on Unix-based operating systems (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].

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.

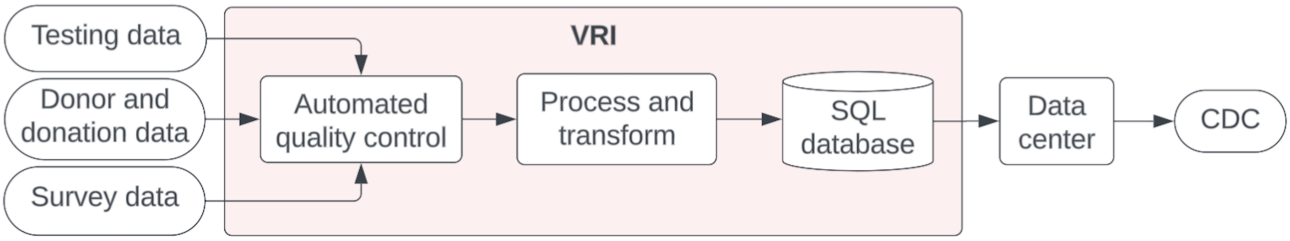
Data Dictionary

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, 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 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).

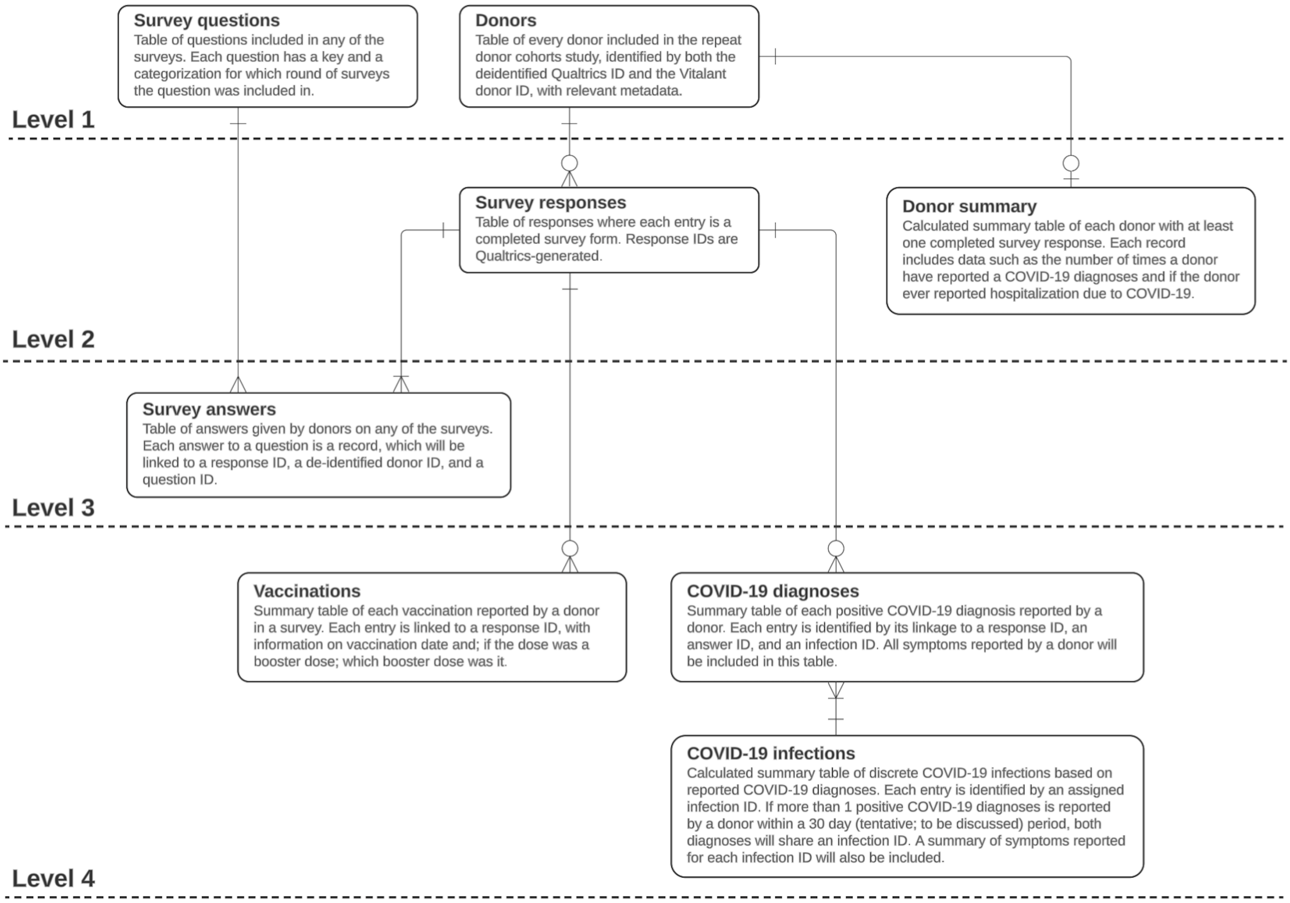
Laboratory Testing Data

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.



Survey questions
Table of questions included in any of the surveys. Each question has a key and a categorization for which round of surveys the question was included in.

Donors
Table of every donor included in the repeat donor cohorts study, identified by both the deidentified Qualtrics ID and the Vitalant donor ID, with relevant metadata.

Survey responses
Table of responses where each entry is a completed survey form. Response IDs are Qualtrics-generated.

Donor summary
Calculated summary table of each donor with at least one completed survey response. Each record includes data such as the number of times a donor have reported a COVID-19 diagnoses and if the donor ever reported hospitalization due to COVID-19.

Survey answers
Table of answers given by donors on any of the surveys. Each answer to a question is a record, which will be linked to a response ID, a de-identified donor ID, and a question ID.

Vaccinations
Summary table of each vaccination reported by a donor in a survey. Each entry is linked to a response ID, with information on vaccination date and; if the dose was a booster dose; which booster dose was it.

COVID-19 diagnoses
Summary table of each positive COVID-19 diagnosis reported by a donor. Each entry is identified by its linkage to a response ID, an answer ID, and an infection ID. All symptoms reported by a donor will be included in this table.

COVID-19 infections
Calculated summary table of discrete COVID-19 infections based on reported COVID-19 diagnoses. Each entry is identified by an assigned infection ID. If more than 1 positive COVID-19 diagnoses is reported by a donor within a 30 day (tentative; to be discussed) period, both diagnoses will share an infection ID. A summary of symptoms reported for each infection ID will also be included.

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.

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.

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.

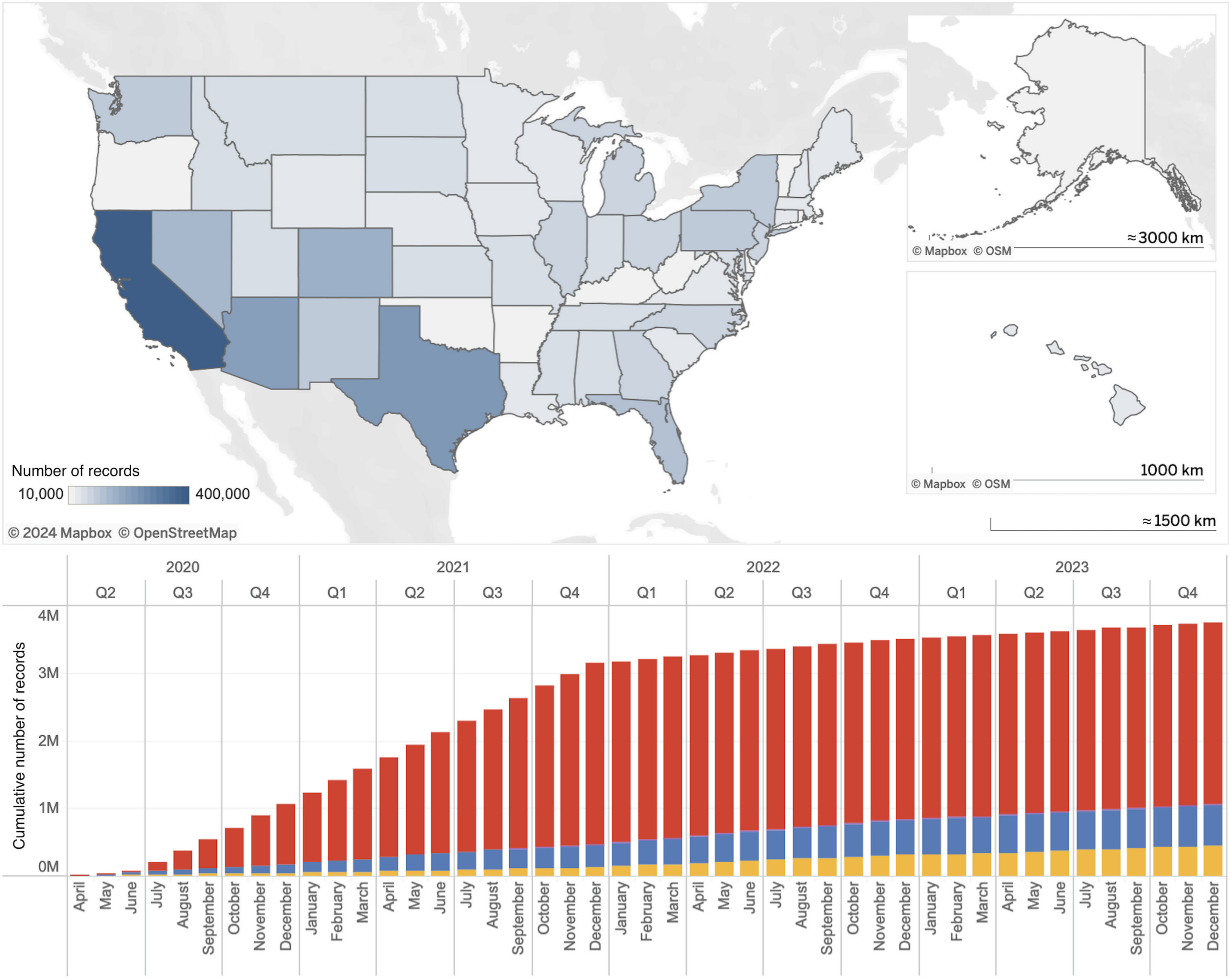
Evaluation of Use Case

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.

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.

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, 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

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]

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

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].

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 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.

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.

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 ($\chi^2_{11}=17.612$; $P >.05$). 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.

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 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).

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	

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 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) 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

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]

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

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].

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 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

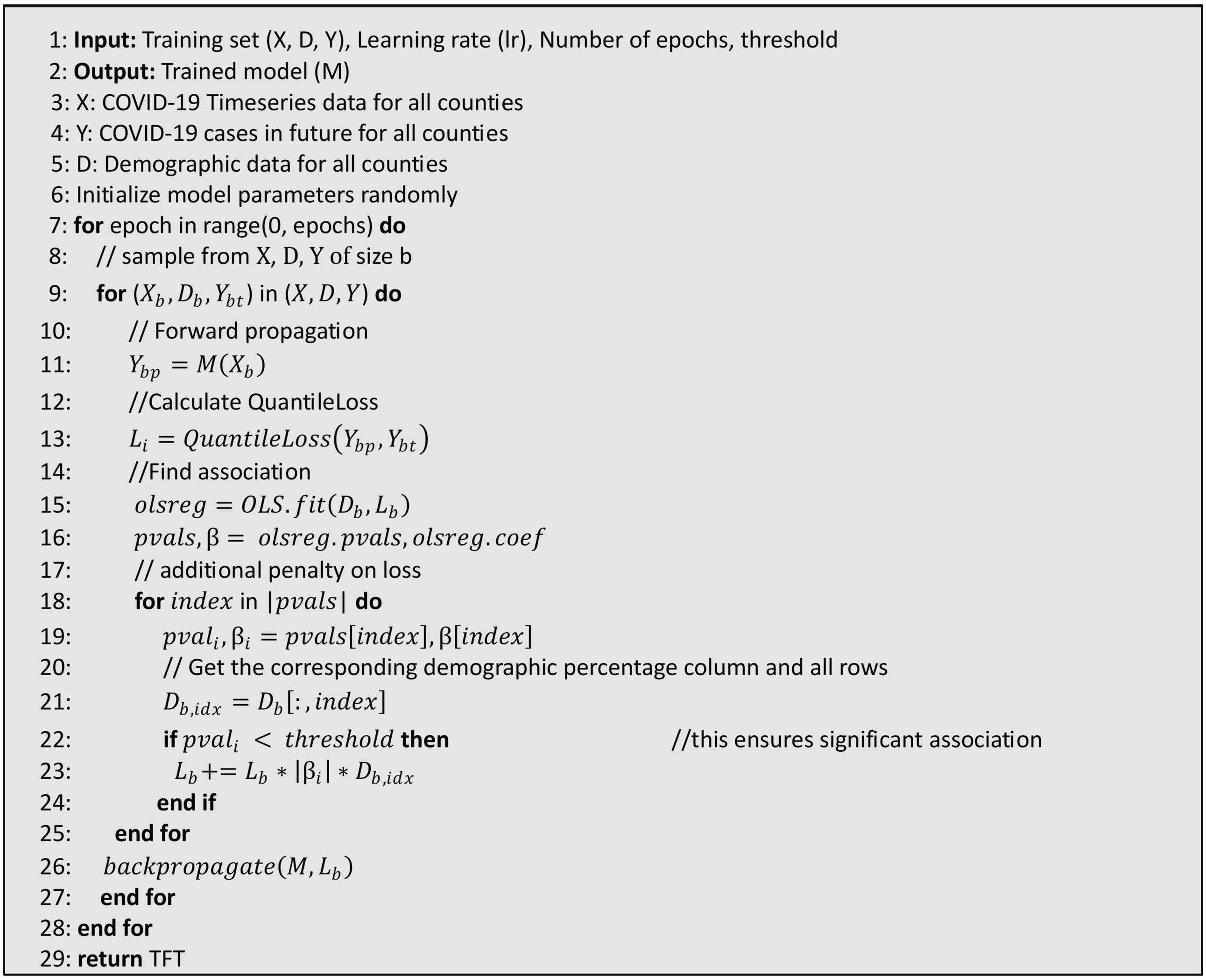
Proposed DemOpts

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.

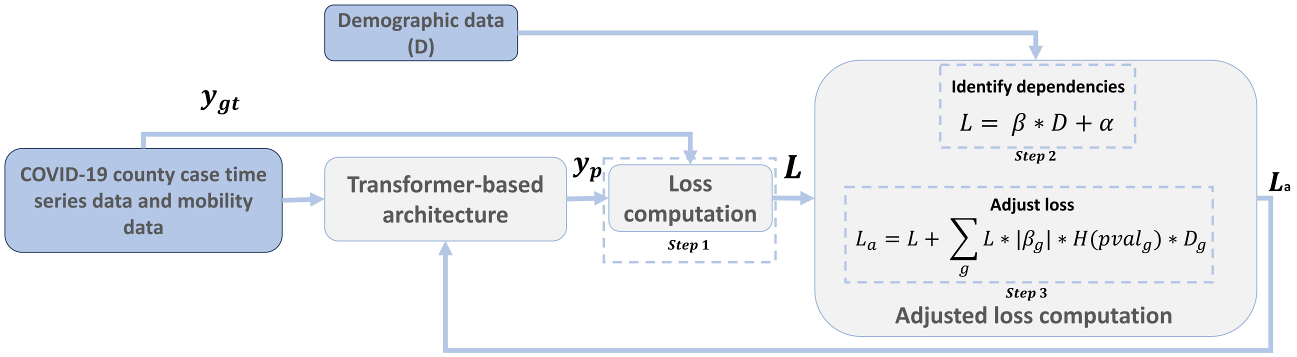
The algorithm can be divided into 3 steps (refer to Figure 1, Figure 2, and "S.1 DemOpts Method" in Multimedia Appendix 1 for mathematical details).

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
        $olsreg = OLS.fit(D_b, L_b)$
        $pvals, \beta = olsreg.pvals, olsreg.coef$
        // additional penalty on loss
        for $index$ in $|pvals|$ do
            $pval_i, \beta_i = pvals[index], \beta[index]$
            // Get the corresponding demographic percentage column and all rows
            $D_{b,idx} = D_b[:, index]$
            if $pval_i < threshold$ then                //this ensures significant association
                $L_b += L_b * |\beta_i| * D_{b,idx}$
            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 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:

$$Ladj=PBL(yip,yi)+\sum jH(pvalj)(|\beta j|*Dj*L) \quad \text{where}$$

$$H(x)=\{1 \quad if x<0.0050 \quad if x\geq 0.005$$

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 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.

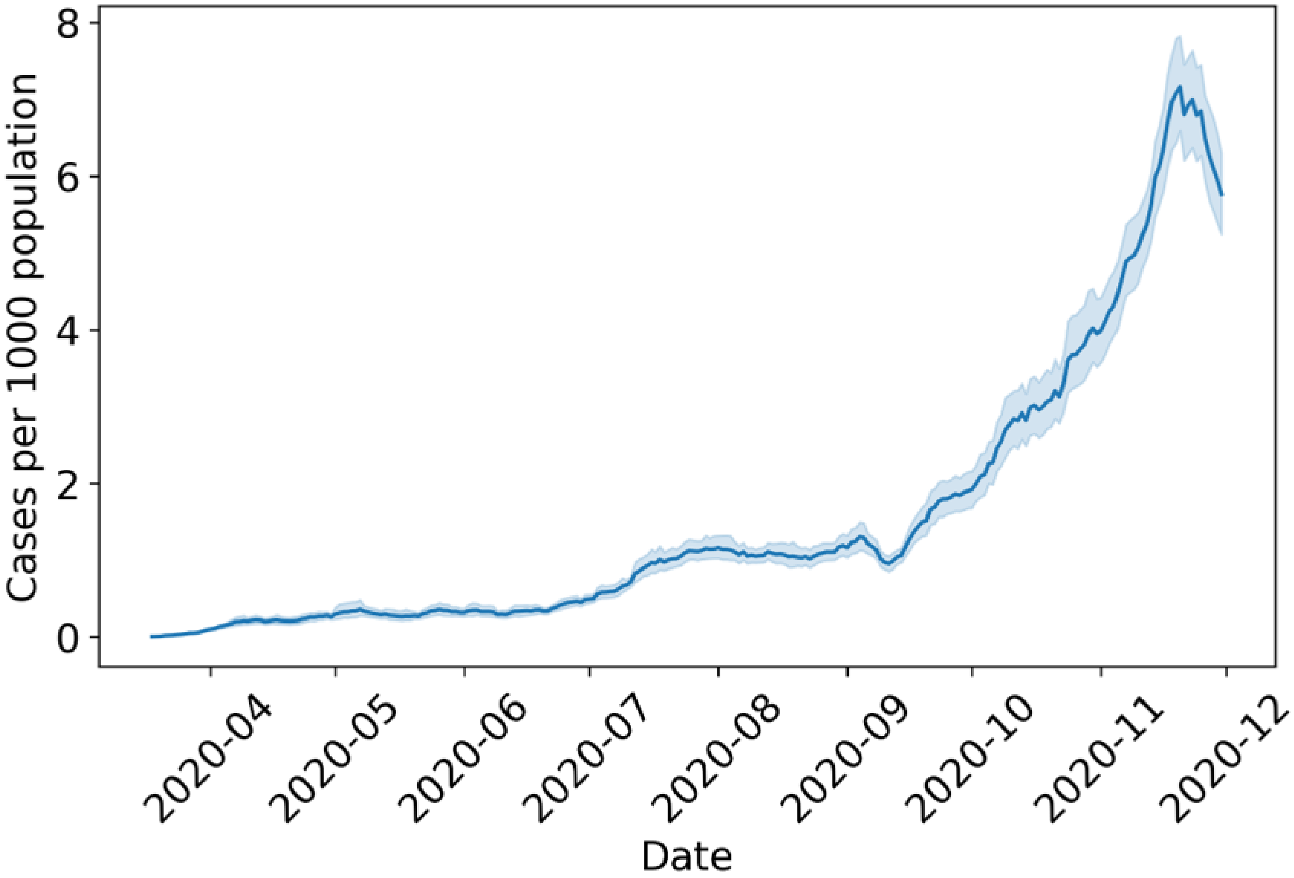
Datasets

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 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.

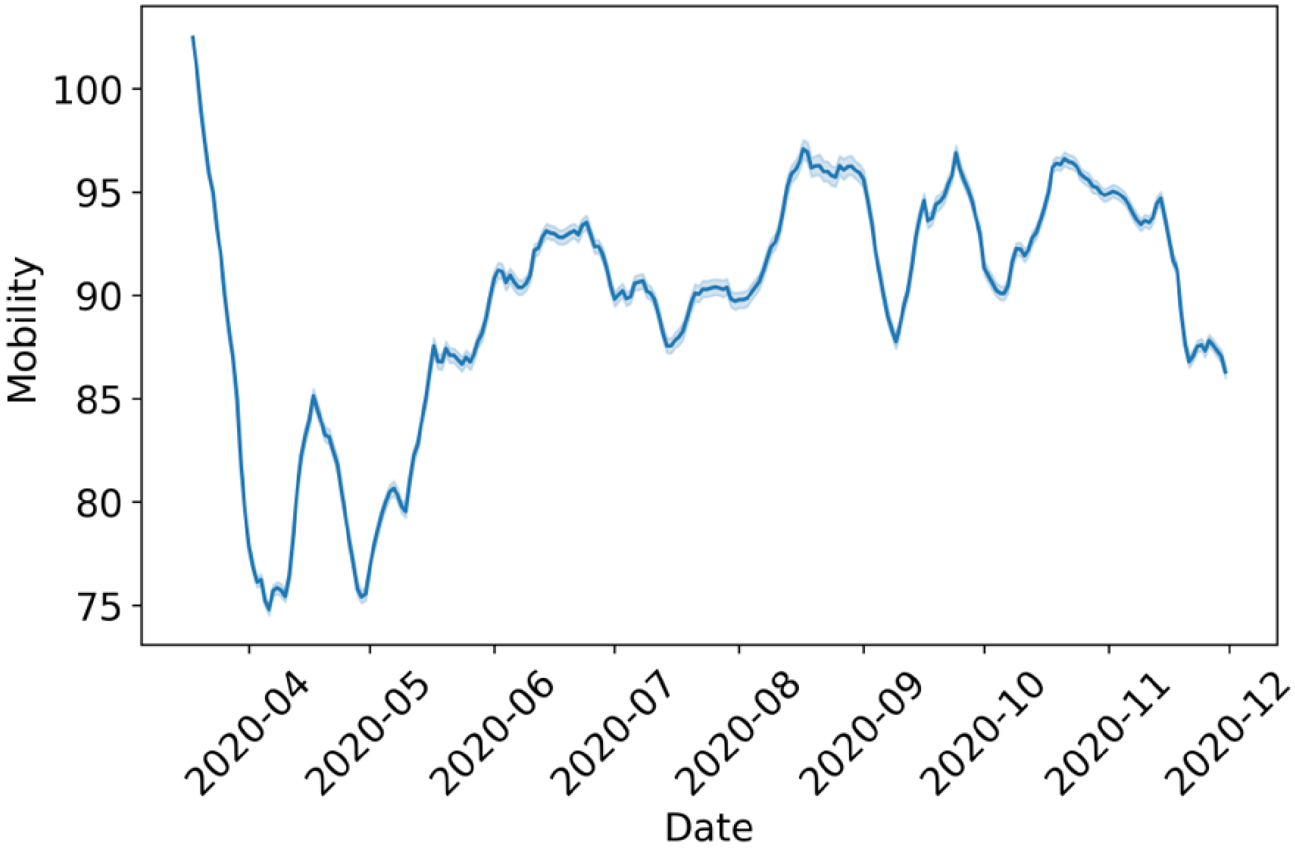


Mobility Data

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.

Figure 4. Mobility for all ethnic and racial groups.



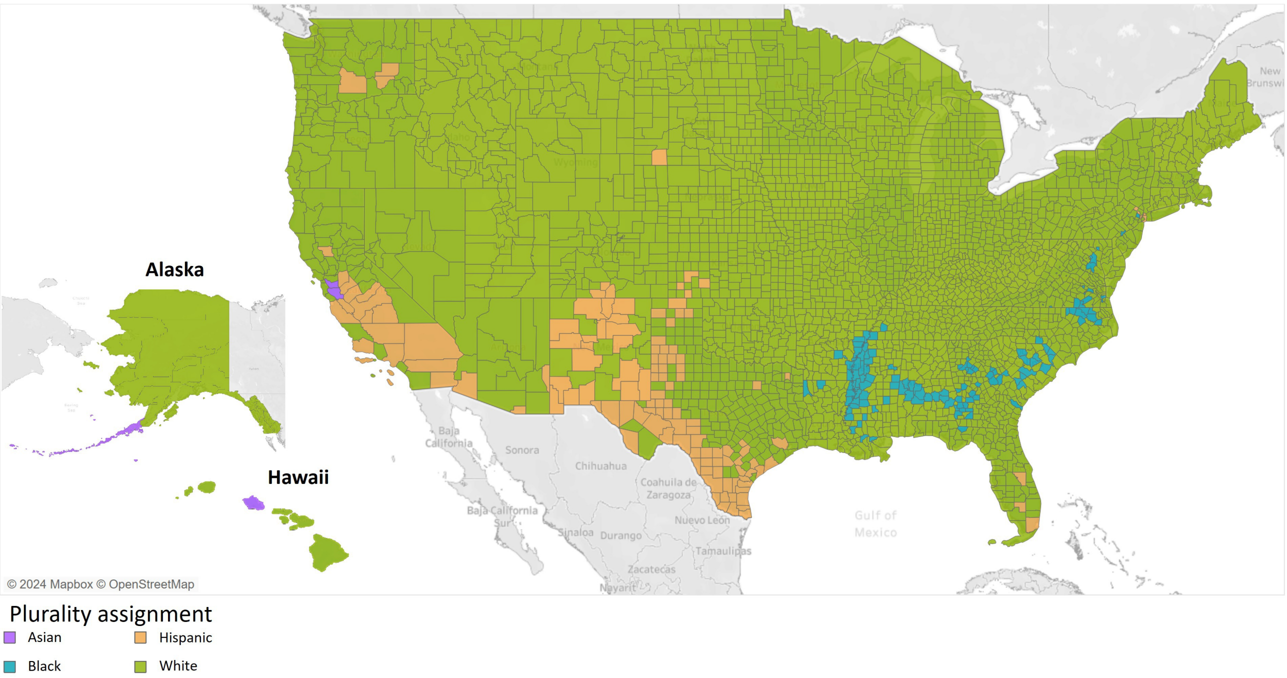
Race and Ethnicity Data

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)

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 $\text{TFT}_{\text{Baseline}}$, a TFT prediction model without a debiasing method; (2) the $\text{TFT}_{\text{Individual}}$, $\text{TFT}_{\text{Group}}$, and $\text{TFT}_{\text{Sufficiency}}$, TFT prediction models with state-of-the-art debiasing methods; and (3) $\text{TFT}_{\text{DemOpts}}$, 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 $\text{TFT}_{\text{Baseline}}$ 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 ($\text{NormPBL}_{\text{Group1}}$ - $\text{NormPBL}_{\text{Group2}}$). 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 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

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 $\text{TFT}_{\text{Baseline}}$ (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 $\text{TFT}_{\text{Baseline}}$, which is not enhanced with any debiasing method.

Table . Soft error parity analysis. Each value represents the distance ($|1\text{-AER}_{\text{race}}|$) for each protected group and debiasing method. $\text{TFT}_{\text{DemOpts}}$ 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 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).

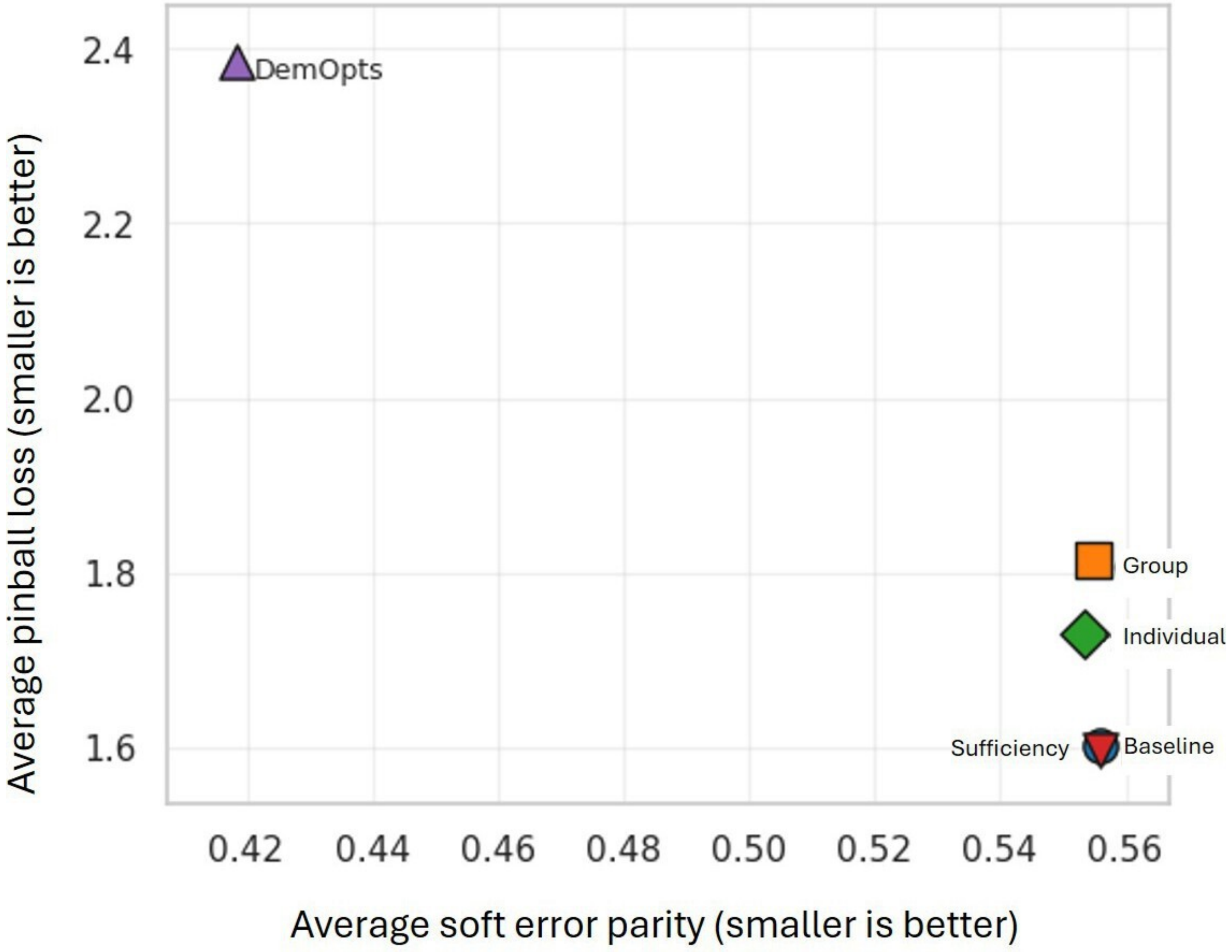
Table . Group-wise pinball loss for each model. Demographic Optimization (DemOpts) has higher average pinball loss compared to the other models. The fairness-accuracy tradeoff leads to slightly larger pinball loss values for DemOpts compared to other methods.

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.



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]

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]

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]

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]

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

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org