Online Journal of Public Health Informatics

High-quality research and innovation in the field of public health informatics
Volume 17 (2025) ISSN 1947-2579 Editor in Chief: Edward K. Mensah, PhD, MPhil

Contents

Viewpoints

Participatory Animation for Health Promotion in Digital-Based Health Interventions: Viewpoint on Methodology and Application (e72737)
Essence Wilson. . . . 5

Cool Solutions in Hot Times: The Case for Digital Health in Heatwave Action Plans (e66361)
Maria Loureiro, Neil Jennings, Emma Lawrance, Daniela Ferreira-Santos, Ana Neves. . . . 16

Remote Consultations in England During COVID-19: Challenges in Data Quality, Linkage, and Research Validity (e66672)
Liliana Hidalgo-Padilla, Massar Dabbous, Kristoffer Halvorsrud, Thomas Beaney, Gideon Gideon, Eoin Gogarty, Geva Greenfield, Benedict Hayhoe, Gabriele Kerr, Rosalind Raine, Nirandeep Rehill, Robert Stewart, Fiona Gaughran, Mariana Pinto da Costa. . . . 22

Technology for Mental Health: Reflections on Scope and Future Directions in Institutes of Higher Education in India (e78065)
Seema Mehrotra, TK Srikanth, Neha Dahiya, Pulkit Verma, Girish Rao, Prachi Sanghvi, Ashoo Grover, Rajesh Sagar. . . . 31

Reviews

Imposters, Bots, and Other Threats to Data Integrity in Online Research: Scoping Review of the Literature and Recommendations for Best Practices (e70926)
Isabella Strickland, Amy Ferketich, Alayna Tackett, Joanne Patterson, Nicholas Breitborde, Jade Davis, Megan Roberts. . . . 46

Reporting of Fairness Metrics in Clinical Risk Prediction Models Used for Precision Health: Scoping Review (e66598)
Lillian Rountree, Yi-Ting Lin, Chuyu Liu, Maxwell Salvatore, Andrew Admon, Brahmajee Nallamothu, Karandeep Singh, Anirban Basu, Fan Bu, Bhramar Mukherjee. . . . 60

Cultivating Disaster Preparedness: Scoping Review of Technology's Contribution to Situational Awareness and Disaster Mindset in Disaster Medicine (e75404)
Amir Khorram-Manesh, Marius Johannessen, Laurits Nielsen, Eric Carlström, Lasse Berntzen, Lene Sandberg, Jarle Sørensen. . . . 382

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Accuracy of Nutrition-Related Awareness Messages on Twitter (Rebranded as X) by the Nutrition Awareness Providers in the Kingdom of Saudi Arabia: Validity Content Analysis (e85577)
Duaa Alammari, Iman Bindayel, Noura Althukair, Noura AlRomi, Khalid Aldubayan, Najla Khateeb, Ahmed Alabdrabalnabi, Banan Banamah. 7 1

Original Papers

Bayesian Network Analysis of Intervention-Induced Physical Activity Behavior Change: Comparative Modeling Study Across Age, Education, and Activity Impairment Subgroups (e57977)
Simone Tummers, Arjen Hommersom, Lilian Lechner, Roger Bemelmans, Catherine Bolman. 73

Reconsidering Trust and Information Engagement and Unpacking the Role of Emotion in Public Responses During the Early Stage of a Public Health Crisis in China: Web-Based Survey Study (e77790)
Zhiming Liu, Jiawei Tu, Tien-Tsung Lee, Lu Wei. 93

Evidence-Based Mental Health at Scale: Benchmarking Retrospective Cohort Study of a Digital Employee Benefits Program for Depression and Anxiety (e72999)
Emily Ward, Matt Hawrilenko, Millard Brown, Adam Chekroud. 107

Application of Machine Learning to Auto-Code Injury Data in the e-CHIRPP System: Development and Evaluation Study (e69143)
Shamir Mukhi, Steven McFaull, Wendy Thompson, Tim Beattie. 127

Forecasting Neonatal Mortality in Ethiopia to Assess Progress Toward National and International Reduction Targets Using Classical Techniques and Deep Learning: Time-Series Forecasting Study (e66798)
Shimels Kebede. 136

Understanding Patient Perceptions of Bacterial Vaginosis Treatments: Mixed Methods Sentiment Analysis Study of Online Drug Review Forums (e71720)
Eren Watkins, Andrea Cimino, Curtis Culbertson, Juliana Raymaker, Jennifer Amico. 148

Quantifying Patient Demand for Orthopedics Care by Region Through Google Trends Analysis: Descriptive Epidemiology Study (e63560)
Abram Qiu, Kristopher Meadows, Fei Ye, Osasu Iyawe, Kenneth Kenneth-Nwosa. 162

The Impact of the Burden of COVID-19 Regulatory Reporting in a Small Independent Hospital and a Large Network Hospital: Comparative Mixed Methods Study (e63681)
Yalini Senathirajah, David Kaufman, Kenrick Cato, Pia Daniel, Patricia Roblin, Andre Kushniruk, Elizabeth Borycki, Emanuel Feld, Poli Debi. . 1 7 1

Use of Biometrics for Records Deduplication: Case Study of the National Data Repository in Nigeria (e67580)
Ademola Oladipo, Ibrahim Dalhatu, Stephen Balogun, Moyosola Bamidele, Ayodele Fagbemi, Isah Abbas, Nannim Nalda, Richard Ugbena, Jude Orjih, Timothy Efuntoye, Brooke Doman, Sadhna Patel, Herman Tolentino, Daniel Rosen, James Kariuki, Johnson Alonge, Kehinde Balogun, Nnamdi Umeh, Gibril Gomez, Oludare Onimode, Olaposi Olatoregun, Jay Samuels, Adebobola Bashorun. 185

Open Source Web Application (HealthTest) for Emotional Health and Wellness Management in University Students: Development and Usability Study (e69413)
Lucrecia Llerena, Daisy Nata Castro, Nancy Rodriguez, Donald Silva Sánchez. 195

XSL•FO
RenderX

Identifying Substance Use and High-Risk Sexual Behavior Among Sexual and Gender Minority Youth by Using Mobile Phone Data: Development and Validation Study (e68013)
Mehrab Beikzadeh, Ian Holloway, Kimmo Kärkkäinen, Chenglin Hong, Cory Cascalheira, Elizabeth Wu, Callisto Boka, Alexandra Avendaño, Elizabeth Yonko, Majid Sarrafzadeh. . . . 205

Usage of Telehealth and Telenutrition Services by Registered Dietitian Nutritionists: Cross-Sectional Study (e80211)
Najlaa Al-Mana, Suhair Abdalla, Asrar Qari, Mohamed Ahmed, Wejdan Alshehri, Lujain Baabdullah. . . . 222

E-Screening for Prenatal Depression in Kampala, Uganda Using the Edinburgh Postnatal Depression Scale: Survey Results (e51602)
Hasifah Namatovu, Mark Magumba, Dickens Akena. . . . 233

Real-World Data on Alcohol Consumption Behavior Among Smartphone Health Care App Users in Japan: Retrospective Study (e57084)
Kana Eguchi, Takeaki Kubota, Tomoyoshi Koyanagi, Manabu Muto. . . . 244

Accuracy of Nutrition-Related Awareness Messages on Twitter (Rebranded as X) by the Nutrition Awareness Providers in the Kingdom of Saudi Arabia: Validity Content Analysis (e68128)
Duaa Alammari, Iman Bindayel, Noura Althukair, Noura AlRomi, Khalid Aldubayan, Najla Khateeb, Ahmed Alabdrabalnabi, Banan Banamah.
2 5 2

Data Dashboard Acceptability, Use, and Perceived Effectiveness in Disseminating Local Overdose Data and Resources in a Rural New York State County: A Cross-Sectional Study (e68977)
Zhongxuan He, Monika Salvage, Corinna Noel. . . . 267

Posttraumatic Stress Disorder Content on TikTok: Cross-Sectional Analysis of Popular #PTSD Posts (e71209)
Brittany Rohl, Laura Jones, Rachel Nattis, Robert Claar, Xavier Velez, Joy Gabrielli, John Williamson, Eric Porges. . . . 280

Expansion and Assessment of a Web-Based 24-Hour Dietary Recall Tool, Foodbook24, for Use Among Diverse Populations Living in Ireland: Comparative Analysis (e52380)
Grace Bennett, Shuhua Yang, Laura Bardon, Claire Timon, Eileen Gibney. . . . 291

One Health Index Calculator for India Using Empirical Methods for Policy Stewardship: Development and Usability Study (e65039)
Saveetha Meganathan, Arpit Katiyar, Esha Srivastava, Rakesh Mishra. . . . 316

Issues in Identifying Strategies for Youth Mental Well-Being in Stockholm Municipalities Using Participatory Sessions and Text Mining: Qualitative Study (e66377)
Harsha Krishna, Adam Darwich, Sebastiaan Meijer. . . . 326

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 (e63120)
Kaori Yamaguchi, Takemi Akahane, Emi Yasuda, Manabu Akahane. . . . 337

Nowcasting to Monitor Real-Time Mpox Trends During the 2022 Outbreak in New York City: Evaluation Using Reportable Disease Data Stratified by Race or Ethnicity (e56495)
Rebecca Rohrer, Allegra Wilson, Jennifer Baumgartner, Nicole Burton, Ray Ortiz, Alan Dorsinville, Lucretia Jones, Sharon Greene. . . . 353

Evaluation of a Sentinel Hypertension Surveillance System in Mojo, East Shewa Zone, Oromia, Ethiopia: Concurrently Embedded Mixed Design Study (e72909)
Abiyie Demelash Gashe, Yeshiwas Ayale Ferede, Dawit Zenebe Weldemichael, Aman Yesuf Endries. . . . 371

XSL•FO
RenderX

Research Letter

Disparities in Tobacco Use, Binge Drinking, and Injection Drug Use Among Sexual and Gender Minoritized People Recruited Online in San Francisco, California: Survey Results (e81982)
Sean Arayasirikul, Jarett Maycott, Arianna Contestable. 310

XSL•FO
RenderX

Participatory Animation for Health Promotion in Digital-Based Health Interventions: Viewpoint on Methodology and Application

Essence Lynn Wilson, PhD

Department of Psychology, California State University, Northridge, 18111 Nordhoff Street, Northridge, CA, United States

Corresponding Author:
Essence Lynn Wilson, PhD
Department of Psychology, California State University, Northridge, 18111 Nordhoff Street, Northridge, CA, United States

Abstract

Digital-based health interventions (DHIs), defined as health services delivered electronically, have demonstrated effectiveness in promoting health outcomes. However, DHIs often suffer from low user retention, a challenge attributed to limited attention to sociocultural determinants and insufficient user engagement strategies. This paper explores participatory animation (PA), a collaborative methodology that engages community partners in co-creating animated content, as a strategy to improve DHI retention and effectiveness. Drawing from existing literature, this viewpoint examines the theoretical foundations and practical affordances of PA for enhancing DHIs. We describe PA as a multistep production process that integrates participant-driven oral and visual design contributions into multimedia outputs for use in DHIs. Here, PA shows promise in producing engaging and culturally resonant content, with the potential to improve intervention uptake and sustain user engagement. Despite these affordances, PA remains underused in health research. Given the growing urgency to develop effective, equitable DHIs, PA offers a novel, community-informed approach for enhancing both design and implementation. This paper positions PA as a methodological frontier for DHI science.

(Online J Public Health Inform 2025;17:e72737) doi:10.2196/72737

KEYWORDS

digital-based health interventions; participatory methods; health determinants; animation; community-engaged research

Introduction

Health scholars increasingly use digital-based health interventions (DHIs) to improve patient health outcomes through interactive, scalable formats [1]. These interventions encompass digitized health services such as virtual environments and personalized video content, designed to engage diverse populations and address complex health issues [2-4]. Indeed, DHIs have demonstrated effectiveness in increasing treatment uptake, such as for mental health and substance use cessation services [5,6].

Despite their widespread adoption, DHIs' efforts in sustaining participant use have historically underperformed [7-9]. While many interventions demonstrate efficacy in improving health outcomes, often grounded in behavioral science frameworks [6,10], long-term participant retention remains limited [7-9]. Systematic reviews and meta-analyses across eHealth, mobile health, and broader DHIs consistently report moderate to small effect sizes, an outcome posited to implementation shortcomings and low user engagement efforts and user engagement [11-13].

The limited impact of many DHIs highlights gaps in their development, particularly in their ability to address the complex, multilevel determinants of health, often rooted in the interplay between physical, cultural, economic, and political environments [14,15]. However, many DHIs inadequately engaged with these broader influences, limiting their relevance to the target population they intend to serve target population [16]. As such, DHIs are experiencing low participant engagement, as reflected in consistently high attrition and low adherence rates [17-21].

Integrating participatory frameworks in DHI development offers a strategic means to address complex health determinants and improve intervention effectiveness. By fostering equitable collaboration between researchers and the communities they intend to serve, these practices democratize the research process and shift focus beyond individual behaviors to address broader determinants of health, a key consideration in DHI development [22]. Participatory approaches have demonstrated success in enhancing user engagement and retention, particularly within underserved populations [23-25]. They can also empower resource-limited communities by reducing barriers to achieving a healthier environment, making them especially valuable for addressing health disparities in historically marginalized groups [26].

Participatory animation (PA), situated within broader participatory research traditions (eg, written, observational, and visual), offers a distinct method for exploring complex health determinants and enhancing engagement with DHIs [12]. Drawing from the foundational work of Lomax and Smith [27], PA functions as both a methodology and cultural practice that seeks to illuminate the inherent and relational dialogical

XSL•FO
RenderX

processes involved in the co-production of animation. Animation is defined as "any application that generates a series of frames, with each frame appearing as a modification of the previous one," with each sequence or narrative determined by the designer(s) [28,29]. Evidence suggests that animation, when integrated into DHIs, can improve intervention outcomes, including increased patient knowledge [30] and more positive health attitudes and behaviors [31]. Moreover, animated outputs can be iteratively refined into evidence-based interventions, offering a degree of narrative flexibility and substantive participant representation that may not be achievable through other visual or digital modalities.

Drawing from available scholarship, this viewpoint considers PA a promising health promotion method. More specifically, this paper offers a perspective on PA as an emerging methodology for enhancing DHI outcomes, synthesizing theoretical foundations, reflecting on its practical applications, and proposing directions for future research. This paper substantiates DHI science in offering a unique and dualistic strategy for assessing complex health determinants and enhancing intervention features necessary for heightened patient prohealth activity. We discuss this method's features, including context on when to use it, highlight core theoretical frameworks that support its ability to enhance DHI structures, and discuss logistical challenges to ensure transparency in research application. An innovative methodological intervention in health research, we ask the following research questions:

1. What is PA, and what are its unique affordances, including its theoretical foundations, for promoting prohealth outcomes within a digital context?
2. In addition to examining its affordances for health promotion, what are the potential challenges of using PA in health research?

Literature Review

Overview

Healthy People 2030's aim, in line with society's growing reliance on digital resources, is to "make electronic health information easy to understand and use [as this] is key to improving health and well-being" [32]. This proposal has two aims: (1) to encourage robust production of digital resources (eg, DHIs) to improve health and well-being, and (2) to use digital-based mediums that match the target population's existing behaviors and attitudes to its technological capabilities, focusing on the affordances of animation for this paper. Here, we address these claims in the review that follows.

Digital-Based Health Interventions

Emerging technologies offer substantial potential for expanding access to evidence-based health care, particularly through DHIs. These interventions, which are health services delivered through both formal and informal digital platforms, have been increasingly adopted across a variety of health domains. Research consistently shows that DHIs often outperform standard treatments in improving health outcomes [2]. Meta-analytic evidence supports their effectiveness in several areas, including mental health (Hedge G=0.40) [6], physical exercise promotion (mean difference=49) [33], and substance use and smoking cessation (log RR=0.77) [5].

Recent research highlights the potential of DHIs to influence both health behaviors and outcomes. One key advantage is their capacity to reach a broad and often hard-to-reach populations, including those with limited access to traditional health care services [34]. Another advantage is the low-cost per user model, which allows health care professionals and advocates to extend their reach and increase public access to care. Additionally, the private and user-directed nature of DHIs can reduce stigma-related barriers by allowing individuals to engage with health content discreetly, outside public or clinical settings [10].

While DHIs show promise in reducing health disparities by expanding their access to timely, technology-enabled care, their effectiveness is often undermined by poor user retention. High attrition rates remain a consistent challenge across DHI modalities [11-13,20]. For example, Kampmeijer et al's systematic review of eHealth and mobile health interventions [13] among older adults reported that 25% of studies identified motivation-related barriers, including difficulty using new technologies. These retention challenges are frequently attributed to a limited understanding of users' needs, including their technological comfort, motivational drivers, and contextual barriers [35]

Assessing Complex Health Determinants

Enhancing user engagement in DHIs requires attention to factors beyond individual behaviors. The Bronfenbrenner socioecological model (SEM) offers a useful framework for examining multilevel influences that shape health outcomes, particularly among communities facing intersecting structural challenges. The SEM includes 4 primary levels: the individual level, which contains physical, cognitive, and emotional health factors; the interpersonal level, involving social networks and relationships [36,37]; the community level, which considers access to and navigation of institutions such as health care, higher education, and media; and the institutional or structural level, which addresses systemic forces such as economic conditions, political environments, and cultural norms [38]. By incorporating these interconnected layers, the SEM enables a more holistic approach to designing DHIs that resonate with users' lived experiences and contexts.

Empirical applications of SEM demonstrate its use in shaping effective health strategies. For example, Byrd and McKinney [39] applied this model to guide clinical outreach and research initiatives that supported more inclusive campus climates and policies for university students. Building on this, Wilson et al [40] conducted a cross-sectional analysis to address mental health disparities among African American college students. Their findings highlighted the need for institutional policies that recognize race-based discrimination as a significant stressor and emphasized the importance of broader support systems, such as cultural affinity groups and strength-based evaluations of campus health services.

Applying a multilevel framework also informs the design of digital intervention strategies. Fish et al [41] developed Q Chat Space, a digital support platform for LGBTQ+ youth by

XSL•FO
RenderX

accounting for contextual factors such as geographic location and family structures to optimize engagement. Similarly, Macheroni et al [42] assessed a survey of 10,000 European children, finding that perceived individual and social discrimination was negatively associated with self-efficacy in digital skill development. They recommend targeted intervention to address offline discrimination to promote uptake and effective use of digital tools. As we discuss in the "Exploring Participatory Frameworks for Health Promotion" section, participatory approaches, particularly PA, are well-positioned to support this goal by integrating community input and contextually relevant information into intervention design.

User Engagement in DHI Structures

Perski et al [43] define user engagement as the extent of digital use, encompassing metrics such as volume, frequency, duration, and depth of digital interaction. Beyond these behavioral measures, engagement also includes subjective experience, such as attention, interest, and emotional investment, that contribute to sustained intervention use [44,45]. Despite the centrality of engagement to intervention success, the factors that enhance user engagement remain underexplored in DHI research. This gap is reflected in persistently high attrition rates and low adherence across [17-19,21].

Visual salience and contextually relevant design play a critical role in enhancing user engagement with DHIs [23]. Studies consistently show that the perceived attractiveness of an intervention positively influences user retention [46,47]. Among the most effective strategies are the integration of context-congruent animation and digital narratives, both of which have been shown to improve message uptake and promote prohealth behaviors [48-51].

Despite growing interest in DHIs, few interventions are specifically designed to reflect racial or cultural experiences, which can limit engagement among diverse populations [16]. Jones et al [52] identified several engagement-enhancing affordances when DHIs directly addressed racial stressors and incorporated culturally affirming elements, such as messages promoting racial pride. Embedding positive cultural context, such as racial identity and socialization or "positive verbal and non-verbal messages, actions, and beliefs, conveyed about [cultural factor, or in this case race] increasing understanding of the …ecology" can deepen ecological understanding and improve relevance to the target population.

Exploring Participatory Frameworks for Health Promotion

Incorporating a participatory framework into DHI development can help address persistent challenges in user retention (Table S1 in Multimedia Appendix 1). Here, participatory practices can (1) assess formative data and enhance data triangulation to understand the scope of determinants, (2) design intervention components to increase engagement, and (3) expand networks for outreach by ascertaining ecological barriers and facilitators [53]. Specific to digitized spaces, community partners can collaborate among researchers and stakeholders for content design and provide iterative feedback on feasibility and acceptability [54,55]. For example, June and Aguilar's case study of the Central Texas African American Family Support Conference [53] demonstrated how shared understanding and collective responsibilities between community partners and stakeholders fostered long-term engagement and program sustainability. Such frameworks are particularly effective in underserved populations, where socioecological complexity requires grounded, community-informed strategies. Indeed, community involvement, if not immersion, can lead to successful engagement and heighten prohealth efficacy outcomes [23,25].

Despite the potential of participatory approaches in technology-enabled interventions, few DHIs have fully integrated such methods [23]. A notable exception is MindLight, an anxiety prevention video game developed by Schoneveld et al [56] for school-age children. In this project, researchers collaborated directly with children diagnosed with anxiety disorders, co-creating monster character designs that were later incorporated into the game. The participatory design process not only acknowledged the children's lived experiences, but also enhanced the game's ability to capture and retain users. This intervention led to a statistically significant reduction in anxiety symptoms between pre-post test conditions, illustrating the value of embedding participatory elements in DHI development.

In this paper, we define a participatory framework as a methodological approach that centers equitable collaboration between researchers and community stakeholders, emphasizing co-design, iterative feedback, and shared ownership of knowledge production [25,26]. This definition draws from foundational community-based participatory research literature [14,26,57] and human-centered design strategies applied to digital health [54].

Participatory Animation

Overview

This paper gives a perspective on PA as a unique opportunity to break down barriers in traditional health research. It highlights its potential as an innovative approach in health promotion geared towards comprehensive intervention design and implementation. Its future directions include refining intervention strategies based on community feedback to sustain engagement and enhance health outcomes.

More specifically, this work is best understood as a conceptually grounded synthesis, integrating existing literature with illustrative examples drawn from prior studies. While no new primary data were collected, the analysis was informed by participatory design principles and grounded in empirical case applications from the field.

Affordances

PA is a methodological and cultural practice centered on the co-creation of animated content [27]. This approach has two key affordances: (1) the co-production process illuminates inherent relational and dialogical processes between participants and researchers, and (2) the resulting animations can serve as multimedia tools to enhance DHI outcomes [48,58]. Indeed, animation as a participatory practice has been effective for

XSL•FO
RenderX

digital health promotion, including areas in community wellness [59], building healthy interpersonal connections [60,61], developing culturally responsive therapeutic practices [62], and fostering safety and belonging for refugee children [63]

Animation as a medium illustrates a unique potential for community engagement. It is defined as "any application that generates a series of frames, with each frame appearing as a modification of the previous one," where each sequence or narrative is determined by the designers [28,29]. Animation can break down the barriers of written text, making content accessible, and can be culturally tailored to reflect the lived experiences of specific target populations. Its iterative nature also allows for ongoing input and refinement, enabling participants to co-construct meaning through a combination of oral and visual storytelling.

The development of animated content can foster civic engagement by empowering community partners to share their experiences, perspectives, and ideas. PA brings together diverse forms of experiences, knowledge, and skills, creating opportunities to explore multilevel health determinants through collaborative design [26,57,64,65]. The production process integrates various artistic and linguistic forms, including creative writing through script development, sound and music composition, and visual design such as character modeling and background art. The multidimensional input not only amplifies participant voices but also enables researchers to gather intersecting forms of data, thereby enhancing the richness of analysis and opportunities for data triangulation.

PA can semiotically manifest an individual's experience by empowering community partners to shape their self-presentation, offering narrative control [27]. Animation production can democratize the research process, countering traditional models in which participant voices are often filtered through external interpretation [66,67]. By fostering collaboration and mutual respect, PA can also serve as a mechanism for building trust between community partners and researchers [68]. For example, Lomax and Smith [27] used animation to capture children's emotional experience during the COVID-19 lockdown. The use of participant-generated voiceovers preserved nuance, such as humor and tone, not easily conveyed through transcripts. Similarly, Massijeh [63] used animation as a tool for play and exploration with South Sudanese refugee children, enabling them to articulate ideas that may have been inaccessible through traditional methods. These examples highlight how PA can enrich data collection and enhance the relevance of formative assessment [69].

Indeed, participatory practices can help address the limitations of traditional qualitative methods like focus groups and interviews. Animation can enhance communication by capturing emotional nuance, symbolic meaning, and complex messages that may be difficult to articulate through verbal methods alone. It can also overcome language barriers, offering a meta-narrative format that converts ideas visually and universally. Vaughn and Jacquez [15] demonstrated this effectively in their community-based project with youth partners aimed at reducing violence and promoting neighborhood well-being. Their iterative process, involving story vignettes, storyboarding, voiceover recordings, and animated revisions, revealed facilitators that may have been overlooked through conventional research, such as the partners' preference for visually appealing, short-form messaging tailored to peer audiences.

PA can be used to explain causal relationships and pathways, whether through multilingual assessments or empirical testing of its messaging effects. Offering multiple textual domains—visual, auditory, and narrative—this approach enables the exploration of complex social structures that may be inaccessible through more traditional or unidimensional methods. For instance, Janssen et al [70] documented how employees used animation to simulate organizational processes and support the adoption of shared services. Their participatory process created a direct pipeline from data collection to content creation, culminating in a film presented to stakeholders for evaluation. As a determinant of efficiency in the business setting, animation, in this case, is used as a participatory praxis to assess causal links while providing a product to engage clients.

Finally, PA also serves as an effective and timely tool for disseminating health messages to diverse community segments. Research indicates that animation can enhance DHI structures and enhance messaging, particularly through widely used digital devices, including smartphones, tablets, and social networking sites [30,31]. Vaughn and Jacquez [15] collaborated with community members to develop violence-prevention animations, which were strategically shared through social networking channels to maximize relevance and reach. This approach illustrates how PA can extend the lifecycle of community-driven interventions beyond the research setting and into everyday digital environments. Here, they co-produced violence-prevention animated content with residents and disseminated it to targeted social networking outlets.

Overall, PA can be understood along 2 key dimensions: as a process and as a product. As a process, this method involves iterative collaboration with community members through processes such as scripting, voiceover development, storyboarding, and visual design, a creative cycle that fosters trust, reflection, and shared meaning-making. As a product, the resulting animated content serves as a multimodal tool for intervention delivery, communication, or knowledge dissemination. This distinction is critical: the process facilitates engagement and data generation, while the product enhances message clarity and reach. Recognizing both dimensions helps researchers design PA projects that are both methodologically rigorous and practically impactful [27,48,59].

It is important to note that PA shares similarities with other participatory visual methods such as photovoice, participatory video, and theater. Scholars and practitioners have used these methods to center community voices, facilitate collective reflection, and aim to shift power, if not democratize, research processes. However, PA offers distinct affordances. Unlike photovoice, which relies on still imagery, or participatory video, which typically requires synchronous participation and access to equipment, animation allows for asynchronous collaboration. Here, community members can contribute ideas, voiceovers, or drawings through diverse and flexible timelines. Compared to theater, which foregrounds live embodiment and performance,

XSL•FO
RenderX

PA can preserve anonymity and privacy, which may be especially valuable in stigmatized health contexts. Moreover, its multimodal nature—blending sound, image, text, and movement—makes it uniquely suited for communicating complex, layered health determinants across digital platforms. Indeed, these affordances specifically position PA as a scalable and context-sensitive approach for DHI development, particularly for and by communities with varying levels of technological and physical access.

Theoretical Foundation

PA is supported by multiple theoretical frameworks that underscore its value for advancing DHI research and practice (Table S2 in Multimedia Appendix 1). From a dialogical framework perspective, animation production unfolds through a series of phases that enable diverse modes of reflection and communication. Indeed, "much of the work [in animation development] occurs under closed doors and is made of incredibly complex intersubjective exchanges, negotiations, and contestations," offering rich opportunities to collect and assess complex health determinants [71]. Within this framework, PA can help address what Bourdieu et al [72] termed the 'paradox of communication' or the difficulty of conveying lived experience within dominant linguistic or symbolic structures. As Frank argues [73], dialogical narratives are especially valuable for disrupting hegemonic frames and making space for alternative or marginalized forms of expression.

The multimodal nature of animations allows for the transmission of rich, layered information shaped by its creators. Frank's dialogical narrative analysis [73] provides a useful lens for understanding this potential, emphasizing the value of examining tensions across different communication modes. His framework identifies 3 distinct publics that help explore the association between visual storytelling and community output evaluation. The first public comprises the core group of participants and researchers or filmmakers, fostering idiosyncratic relationships. The second public includes individuals outside the research project who clearly understand its context and have indirect connections, such as film viewers. The third public, a broader audience, views the film independently and lacks the opportunity for direct engagement or participation in an intersubjective process. This 3-tiered model supports a comprehensive exploration of animation's communicative reach and highlights its potential to foster engagement and meaning-making across multiple levels of audience participation.

Animation alone is supported by numerous cognitive theories that can bolster DHIs' structures. It is supported by the cognitive theory of multimedia learning proposed by Mayer and Moreno [74]. This theory posits that learners engage in three critical cognitive processes when exposed to multimedia content that incites auditory and visual sensations: (1) incoming verbal information to create text schema, (2) incoming visual information to form an image base, and (3) a combination of verbally and visually based models which enhance knowledge acquisition, a variable shown to encourage prohealth behaviors [75]. Cognitive scholars Berney and Bétrancourt [76] found this framework helpful for understanding animation's empirical effects in enhancing participant learning across 50 studies compared to static illustration. These animations simultaneously presented verbal and visual materials that enhanced participant knowledge on the subject.

The social cognitive theory, developed by Bandura [75], further reinforces the value of animation in enhancing DHIs. This theory posits that individuals acquire knowledge not only through direct experience but also by observing others, whether through social interactions or mediated representations. When paired with the cognitive theory of multimedia learning, animations become a powerful tool for modeling behavior, allowing users to visualize actions, consequences, and strategies in ways that support observational learning. O'Day [77] elucidates animation's effects in enhancing modeling structures, where teaching procedural information through dynamic events was more effective than static sequential images.

The cognitive load theory also supports the strategic use of animation in DHIs [78]. This theory emphasizes the limitations of working memory and the importance of minimizing extraneous cognitive demands to facilitate learning and long-term retention. Animation can reduce cognitive load by integrating visual and verbal information in a way that supports comprehension. For instance, animation that includes complementary text content congruent to the targeted message and signaling cues (eg, highlighting, arrows, and flashing symbols) helps direct user attention and reinforce key concepts [79,80]. These features enable users to process complex information more effectively, thereby improving message clarity and enhancing prohealth outcomes [81].

Overall, PA leveraged across numerous research capacities, contingent upon the researcher's epistemological perspective. Researchers may opt for a critical, interpretative, practical, or positivist approach.

Discussion

This paper explored PA as a novel methodological approach for improving DHIs, particularly in addressing persistent issues of low user engagement and retention. Through a synthesis of theoretical frameworks and empirical examples, we identified this method's key affordances, including its potential to incorporate sociocultural determinants, foster participant co-creation, democratize research settings, and enhance message clarity through multimodal storytelling. While underused in the current DHI design, PA offers unique benefits for tailoring interventions to diverse and hard-to-reach populations and increasing their resonance and sustainability.

Effective implementation of PA requires careful consideration of several logistical challenges, particularly around data interpretation. Given the multimodal nature of the medium, there is currently no standardized approach for analyzing its components such as storyboards, voiceovers, or scripts, which can complicate methodological rigor. Existing scholarship has primarily explored PA within the context of health messaging, yet analytic strategies remain highly dependent on the researcher's disciplinary orientation. For example, some may focus on disseminating co-created content through digital platforms to enhance message delivery, while others may treat

XSL•FO
RenderX

the animation process as a data source, applying qualitative methods such as thematic or discourse analysis to examine its narrative and symbolic dimensions. This variability underscores the need for clearer guidance and methodological transparency in future applications [82].

Editing completed animated content for intervention refinement presents several challenges. Most iterative testing tends to occur during the storyboard phases, where collaborators co-develop a visual plan outlining the animation frame-by-frame. However, once production begins, making substantive changes can be labor-intensive and costly. Creating high-quality content, comparable to that produced by commercial studios, can cost up to $7000 per minute [83]. Costs vary depending on how labor is allocated; for example, budgets can be reduced by limited preproduction work, such as concept art, or by using prebuilt assets and lower-frame-rate animation tools. There are also cost-efficient alternatives to producing animation, such as Vyond, PowToon, or Canvas animation, all of which enable users to generate animated presentations with easy-to-use features. Meanwhile, university-community partnerships with student animators, apprenticeship (eg, Legends Animated), or with freelance animators can also offset costs [84].

Nonanimated media may be more appropriate for specific community segments. It is critical to acknowledge regional perspectives on animation when implementing PA. Community partners from European and North American regions may robustly perceive animation as an entertainment source for children [85]. This is reflected in the current PA scholarship, where researchers have predominantly worked with children in these areas. However, Skropanic's study on graphic novels [86], which consists of similar production methods as animation, found this method palpable among older adult veterans to cope with posttraumatic stress disorder symptoms. While research continues to uncover animation's efficacious role in promoting prohealth outcomes, regional perspectives may potentially stifle the implementation of PA among older populations. However, future participatory methods could include co-education or media literacy efforts as part of the rollout phase to normalize animation for adult audiences.

Animation remains widely used in professional settings, often requiring rigorously trained experts to create quality content. As such, it can be time-consuming, taking up to 10 weeks to craft a 60 to 90-second animation. However, the engagement between experts and community partners offers community-building opportunities, particularly with emerging technology. Nonprofit sectors have shown immense community-based engagement practices for animation. One such organization is Legends Animated, a nonprofit that explicitly hires animation apprentices of diverse backgrounds to provide valuable experiential learning and needed client-based artwork for their job portfolio [84]. They foster the capacity to create and work alongside the community sectors and stakeholders to develop animation relevant to their cultural experience. Software like Adobe Animate and Visem could unite diverse talent and voices of varying technological competencies, further democratizing the research setting. It is also worth noting that recent digital animation programs may have a manageable learning curve, making them accessible to creators of varying technical and artistic expertise.

Some ethical and accessibility concerns should also be addressed. To ensure equitable participation from co-creation to dissemination, rotating leadership alongside consent processes, especially as diverse skill sets are leveraged, should be considered [57,66]. This includes potential anonymity in voiceover recordings and visual masking or team names within credits. Specific to accessibility, user-friendly design should be at the forefront of the co-creation process. These include closed captions, audio narration, diverse language translations, screener-compatible descriptions, and diverse device compatibility (eg, phone vs tablet aspect ratio).

Although PA holds promise, empirical evaluations, particularly randomized trials or quantitative effectiveness studies, remain scarce. Future work should evaluate the efficacy of PA through mixed methods or controlled studies to establish robust, generalizable findings. Meanwhile, Wilson, Canty, and Miller's [31] meta-analysis revealed that animation was efficacious in promoting prohealth behaviors and attitudes among children and adults, garnering insights into its methodological potential in health sectors. While PA presents risks in development and implementation, its benefits outweigh the associated costs. This medium, which integrates both dynamic and engaging features, has the potential to have a long-lasting impact and be disseminated across various digital structures, making it multimedia deployable.

Conclusions

PA is a promising strategy for health promotion. This method leverages the animation's nuanced development procedure to procure essential health determinants while enhancing messages. It also empowers communities to participate in their health landscapes. Although animation offers multisensory components functional for participatory praxis, it can be limited due to logistical (eg, cost and time consumption) and technological constraints (competencies). Overall, animation can be the tool to unlock and reveal limitless worlds and multiple truths, where one's imagination is necessary to realize one's civic potential and even acknowledge and address urgent health concerns.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Framework for analyzing participatory animation: process and product dimensions.

XSL•FO
RenderX

[DOCX File, 19 KB - ojphi_v17i1e72737_app1.docx]

References

1. Marent B, Henwood F. Digital health: A sociomaterial approach. Sociol Health Illn 2023 Jan;45(1):37-53. [doi: 10.1111/1467-9566.13538] [Medline: 36031756]
2. Soobiah C, Cooper M, Kishimoto V, et al. Identifying optimal frameworks to implement or evaluate digital health interventions: a scoping review protocol. BMJ Open 2020 Aug 13;10(8):e037643. [doi: 10.1136/bmjopen-2020-037643] [Medline: 32792444]
3. Ahn SJ. Virtual exemplars in health promotion campaigns: heightening perceived risk and involvement to reduce soft drink consumption in young adults. J Media Psychol 2018;30(2):91-103. [doi: 10.1027/1864-1105/a000184]
4. Miller LC, Appleby PR, Christensen JL, et al. Virtual interactive interventions for reducing risky sex: adaptations, integrations, and innovations. In: Noar SM, Harrington NG, editors. eHealth Applications: Promising Strategies for Behavior Change 2012:79-95. [doi: 10.4324/9780203149096]
5. Fang YE, Zhang Z, Wang R, et al. Effectiveness of eHealth smoking cessation interventions: systematic review and meta-analysis. J Med Internet Res 2023 Jul 28;25(1):e45111. [doi: 10.2196/45111] [Medline: 37505802]
6. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079. [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
7. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 1;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
8. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021 Apr 29;8(4):e25847. [doi: 10.2196/25847] [Medline: 33913817]
9. Nicholas J, Boydell K, Christensen H. mHealth in psychiatry: time for methodological change. Evid Based Mental Health 2016 May;19(2):33-34. [doi: 10.1136/eb-2015-102278]
10. Tofighi B, Abrantes A, Stein MD. The role of technology-based interventions for substance use disorders in primary care: a review of the literature. Med Clin North Am 2018 Jul;102(4):715-731. [doi: 10.1016/j.mcna.2018.02.011] [Medline: 29933825]
11. Edwards EA, Lumsden J, Rivas C, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447. [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
12. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36(1):393-415. [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
13. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 5;16 Suppl 5(Suppl 5):290. [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
14. Burdine JN, McLeroy K, Blakely C, Wendel ML, Felix MRJ. Community-based participatory research and community health development. J Prim Prev 2010 Apr;31(1-2):1-7. [doi: 10.1007/s10935-010-0205-9] [Medline: 20143162]
15. Vaughn LM, Jacquez F. Participatory research methods—choice points in the research process. J Particip Res Methods 2020;1(1). [doi: 10.35844/001c.13244]
16. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244. [doi: 10.2196/12244] [Medline: 30563811]
17. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
18. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 6;20(6):e199. [doi: 10.2196/jmir.9275] [Medline: 29875089]
19. Karyotaki E, Riper H, Twisk J, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
20. Terceiro L, Mustafa MI, Hägglund M, Kharko A. Research participants' engagement and retention in digital health interventions research: protocol for mixed methods systematic review. JMIR Res Protoc 2025 Jan 3;14:e65099. [doi: 10.2196/65099] [Medline: 39752662]
21. Du S, Zhao Y. Enhancing digital health interventions for medication adherence: considerations for broader applicability and long-term impact. J Med Internet Res 2025 Mar 14;27:e69204. [doi: 10.2196/69204] [Medline: 40085837]
22. Newton-Lewis T, Nanda P. Problematic problem diagnostics: why digital health interventions for community health workers do not always achieve their desired impact. BMJ Glob Health 2021 Jul;6(Suppl 5):e005942. [doi: 10.1136/bmjgh-2021-005942] [Medline: 34312157]

XSL•FO
RenderX

23. Bergin AD, Vallejos EP, Davies EB, et al. Preventive digital mental health interventions for children and young people: a review of the design and reporting of research. NPJ Digit Med 2020;3(1):133. [doi: 10.1038/s41746-020-00339-7] [Medline: 33083568]
24. Montero M. Community action and research as citizenship construction. Am J Community Psychol 2009 Mar;43(1-2):149-161. [doi: 10.1007/s10464-008-9224-6] [Medline: 19130211]
25. Neuhauser L. Integrating participatory design and health literacy to improve research and interventions. Inf Serv Use 2017 May;37(2):153-176. [doi: 10.3233/ISU-170829]
26. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. Am J Public Health 2010 Apr 1;100 Suppl 1(Suppl 1):S40-S46. [doi: 10.2105/AJPH.2009.184036] [Medline: 20147663]
27. Lomax H, Smith K. Seeing as an act of hearing: making visible children's experiences of the COVID-19 pandemic through participatory animation. Sociol Res Online 2022 Sep;27(3):559-568. [doi: 10.1177/13607804221087276]
28. Bétrancourt M, Tversky B. Effect of computer animation on users' performance: a review. Travail Humain (Paris) 2000;63(4):311-329 [FREE Full text]
29. Maio A. What is animation—definition, history, and types of animation. StudioBinder. 2023 May 7. URL: https://www.studiobinder.com/blog/what-is-animation-definition/ [accessed 2025-12-02]
30. Feeley TH, Keller M, Kayler L. Using animated videos to increase patient knowledge: a meta-analytic review. Health Educ Behav 2023 Apr;50(2):240-249. [doi: 10.1177/10901981221116791] [Medline: 35950326]
31. Wilson EL, Canty S, Miller L. Can animation enhance digital health interventions? a meta-analysis investigating the cognitive components of animation for promoting change. 2023 Presented at: Internation Communication Association; May 25-28, 2023 URL: https://cdn.ymaws.com/www.icahdq.org/resource/resmgr/conference/2023/ICA2023-Program_sm.pdf [accessed 2023-05-27]
32. Diabetes. Office of Disease Prevention and Health Promotion, US Department of Health and Human Services. URL: https://odphp.health.gov/healthypeople/objectives-and-data/browse-objectives/diabetes [accessed 2025-12-02]
33. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719. [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
34. Brewer LC, Fortuna KL, Jones C, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512. [doi: 10.2196/14512] [Medline: 31934874]
35. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
36. Bronfenbrenner U. Toward an experimental ecology of human development. American Psychologist 1977;32(7):513-531. [doi: 10.1037/0003-066X.32.7.513]
37. Friedlander LJ, Reid GJ, Shupak N, Cribbie R. Social support, self-esteem, and stress as predictors of adjustment to university among first-year undergraduates. csd 2007 May;48(3):259-274. [doi: 10.1353/csd.2007.0024]
38. Salihu HM, Wilson RE, King LM, Marty PJ, Whiteman VE. Socio-ecological model as a framework for overcoming barriers and challenges in randomized control trials in minority and underserved communities. Int J MCH AIDS 2015;3(1):85-95. [doi: 10.21106/ijma.42] [Medline: 27621990]
39. Byrd DR, McKinney KJ. Individual, interpersonal, and institutional level factors associated with the mental health of college students. J Am Coll Health 2012;60(3):185-193. [doi: 10.1080/07448481.2011.584334] [Medline: 22420695]
40. Wilson EL, Young LE, Stevens R. Multilevel sources of strength and school-based racial discrimination: competing factors associated with mental health among African American college students. Cultur Divers Ethnic Minor Psychol 2024;31(4):699-710. [doi: 10.1037/cdp0000703]
41. Fish JN, Williams ND, McInroy LB, et al. Q Chat Space: Assessing the feasibility and acceptability of an internet-based support program for LGBTQ youth. Prev Sci 2022 Jan;23(1):130-141. [doi: 10.1007/s11121-021-01291-y] [Medline: 34490582]
42. Mascheroni G, Cino D, Mikuška J, Smahel D. Explaining inequalities in vulnerable children's digital skills: the effect of individual and social discrimination. New Media Soc 2022 Feb;24(2):437-457. [doi: 10.1177/14614448211063184]
43. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
44. Lo Presti L, Testa M, Marino V, Singer P. Engagement in healthcare systems: adopting digital tools for a sustainable approach. Sustainability 2019;11(1):220. [doi: 10.3390/su11010220]
45. Mclaughlin M, Delaney T, Hall A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180. [doi: 10.2196/23180] [Medline: 33605897]
46. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. BMJ Open 2015 Nov 30;5(11):e008581. [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]

XSL•FO
RenderX

47. Pennant ME, Loucas CE, Whittington C, et al. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
48. Calderón JL, Shaheen M, Hays RD, Fleming ES, Norris KC, Baker RS. Improving diabetes health literacy by animation. Diabetes Educ 2014 May;40(3):361-372. [doi: 10.1177/0145721714527518] [Medline: 24676274]
49. Lisetti CL, Wagner E. Mental health promotion with animated characters: exploring issues and potential. Presented at: AAAI Spring Symposium: Emotion, Personality, and Social Behavior; Mar 26-28, 2008; Stanford, California p. 72-79 URL: https://aaai.org/papers/0013-SS08-04-013-mental-health-promotion-with-animated-characters-exploring-issues-and-potential/ [accessed 2025-12-02]
50. Braddock K, Dillard JP. Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. Commun Monogr 2016 Oct;83(4):446-467. [doi: 10.1080/03637751.2015.1128555]
51. Moyer-Gusé E. Toward a theory of entertainment persuasion: explaining the persuasive effects of entertainment-education messages. Commun Theory 2008 Aug;18(3):407-425. [doi: 10.1111/j.1468-2885.2008.00328.x]
52. Jones SCT, Anderson RE, Gaskin-Wasson AL, Sawyer BA, Applewhite K, Metzger IW. From “crib to coffin”: Navigating coping from racism-related stress throughout the lifespan of Black Americans. Am J Orthopsychiatry 2020;90(2):267-282. [doi: 10.1037/ort0000430] [Medline: 32105125]
53. Jung H, Aguilar JP. Everyone’s responsibility: Community partnerships and shared commitment for mental health promotion. Soc Work Ment Health 2016 Sep 2;14(5):474-496. [doi: 10.1080/15332985.2015.1130009]
54. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. Health Promot Pract 2020 Jan;21(1):37-48. [doi: 10.1177/1524839919850557] [Medline: 31131633]
55. Unertl KM, Schaefbauer CL, Campbell TR, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. J Am Med Inform Assoc 2016 Jan;23(1):60-73. [doi: 10.1093/jamia/ocv094] [Medline: 26228766]
56. Schoneveld EA, Malmberg M, Lichtwarck-Aschoff A, Verheijen GP, Engels R, Granic I. A neurofeedback video game (MindLight) to prevent anxiety in children: a randomized controlled trial. Comput Human Behav 2016 Oct;63:321-333. [doi: 10.1016/j.chb.2016.05.005]
57. Israel B, Eng E, Schulz A, Parker E. Introduction to methods in community-based participatory research for health. In: Methods in Community-Based Participatory Research for Health: Jossey-Bass; 2005, Vol. 3:26.
58. Tversky B. Some ways that graphics communicate. In: van den Broek M, Bauer P, Bourg P, editors. Working with Words and Images: Oxford University Press; 2002:57-74.
59. Vaughn NA, Jacoby SF, Williams T, Guerra T, Thomas NA, Richmond TS. Digital animation as a method to disseminate research findings to the community using a community-based participatory approach. Am J Community Psychol 2013 Mar;51(1-2):30-42. [doi: 10.1007/s10464-012-9498-6] [Medline: 22395365]
60. Crawford P. Digital animation as a participatory tool for exploring community visions. Environ Plann B Plann Des 2006 Aug;33(4):481-484. [doi: 10.1068/b33036]
61. Foth M. Participation, animation, design: a tripartite approach to urban community networking. AI & Soc 2010 Aug;25(3):335-343. [doi: 10.1007/s00146-009-0263-9]
62. Ashworth J, Reg H. Animation in therapy: The innovative uses of haptic animation in clinical and community therapeutic practice. Assist Technol 2010;5(1):40-42. [doi: 10.5042/jat.2011.0102]
63. Massijeh M. Tales of watan: animation and myth as a participatory storytelling tool with displaced children [Master’s thesis]. : York University; 2018 URL: https://yorkspace.library.yorku.ca/server/api/core/bitstreams/4bb673d8-7d2d-4c7b-b852-c7a6b997928c/content [accessed 2025-12-02]
64. Coombe CM, Schulz AJ, Guluma L, et al. Enhancing capacity of community-academic partnerships to achieve health equity: results from the CBPR Partnership Academy. Health Promot Pract 2020 Jul;21(4):552-563. [doi: 10.1177/1524839918818830] [Medline: 30596283]
65. Suarez-Balcazar Y. Meaningful engagement in research: community residents as co-creators of knowledge. Am J Community Psychol 2020 Jun;65(3-4):261-271. [doi: 10.1002/ajcp.12414] [Medline: 31907950]
66. Scharff DP, Mathews KJ, Jackson P, Hoffsuemmer J, Martin E, Edwards D. More than Tuskegee: understanding mistrust about research participation. J Health Care Poor Underserved 2010 Aug;21(3):879-897. [doi: 10.1353/hpu.0.0323] [Medline: 20693733]
67. Salmon A. Walking the talk: how participatory interview methods can democratize research. Qual Health Res 2007 Sep;17(7):982-993. [doi: 10.1177/1049732307305250] [Medline: 17724110]
68. Jagosh J, Bush PL, Salsberg J, et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. BMC Public Health 2015 Jul 30;15(1):725. [doi: 10.1186/s12889-015-1949-1] [Medline: 26223523]
69. Rutherford R. Challenges surrounding participatory practice research inanimation: the case of fast affected participants. Animation Studies. 2020 Sep. URL: https://oldjournal.animationstudies.org/jessica-rutherford-challenges-surrounding-participatory-practice-research-in-animation-the-case-of-fasd-affected-participants/ [accessed 2025-12-02]

XSL•FO
RenderX

70. Janssen M, Joha A, Zuurmond A. Simulation and animation for adopting shared services: evaluating and comparing alternative arrangements. Gov Inf Q 2009 Jan;26(1):15-24. [doi: 10.1016/j.giq.2008.08.004]
71. Roy A, Kennelly J, Rowley H, Larkins C. A critical discussion of the use of film in participatory research projects with homeless young people: an analysis based on case examples from England and Canada. Qual Res 2021 Dec;21(6):957-974. [doi: 10.1177/1468794120965374]
72. Bourdieu P. In: Thompson J, Raymond G, Adamson M, editors. Language and Symbolic Power: Harvard University Press; 1993. URL: https://monoskop.org/images/4/43/Bourdieu_Pierre_Language_and_Symbolic_Power_1991.pdf [accessed 2025-12-02]
73. Frank AW. Letting Stories Breathe: A Socio-Narratology: University of Chicago Press; 2010. [doi: 10.7208/chicago/9780226260143.001.0001]
74. Mayer RE, Moreno R. Animation as an aid to multimedia learning. Educ Psychol Rev 2002 Mar;14(1):87-99. [doi: 10.1023/A:1013184611077]
75. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
76. Berney S, Bétrancourt M. Does animation enhance learning? A meta-analysis. Comput Educ 2016 Oct;101:150-167. [doi: 10.1016/j.compedu.2016.06.005]
77. O'Day DH. The value of animations in biology teaching: a study of long-term memory retention. CBE Life Sci Educ 2007;6(3):217-223. [doi: 10.1187/cbe.07-01-0002] [Medline: 17785404]
78. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci 1988 Jun;12(2):257-285. [doi: 10.1016/0364-0213(88)90023-7]
79. Ginns P. Meta-analysis of the modality effect. Learn Instr 2005 Aug;15(4):313-331. [doi: 10.1016/j.learninstruc.2005.07.001]
80. Tabbers H. The Modality of Text in Multimedia Instructions: Refining the Design Guidelines: Erasmus University Rotterdam; 2002.
81. Educational animation: explaining complex ideas with animation. Vyond. 2018 Aug 20. URL: https://vyond.com/resources/how-to-explain-complex-ideas-with-animation/#:~:text=Ani [accessed 2025-12-02]
82. Harrison Ginsberg K, Babbott K, Serlachius A. Exploring participants' experiences of digital health interventions with qualitative methods: guidance for researchers. J Med Internet Res 2024 Nov 28;26:e62761. [doi: 10.2196/62761] [Medline: 39607999]
83. How much does animation cost. Creative Humans. URL: https://www.creativehumans.com/blog/how-much-animation-cost#:~:text=%2D%20%2420%2C000%2D%2450%2C000%20Per%20Minute,incorporate%20more%20advanced%20animation%20effects [accessed 2025-12-02]
84. Legends Animated. URL: https://legendsanimated.org/ [accessed 2025-12-02]
85. Dobson N. Historical Dictionary of Animation and Cartoons: Rowman & Littlefield Publishers; 2020. [doi: 10.5040/9798881822101]
86. Skropanic J. Reanimated: veteran turn PTSD therapy into a comic book, cartoon series. Redding. 2020 Jan. URL: https://www.redding.com/story/news/2020/01/03/reanimated- veteran-turns-ptsd-therapy-into-comic-book-cartoon-series/2751879001/ [accessed 2025-12-02]

Abbreviations

DHI : digital-based health intervention
PA: participatory animation
SEM: socioecological model

Edited by E Mensah; submitted 17.02.25; peer-reviewed by A Liguori, S Chi Lik Au; revised version received 08.08.25; accepted 03.11.25; published 05.12.25.

Please cite as:
Wilson EL
Participatory Animation for Health Promotion in Digital-Based Health Interventions: Viewpoint on Methodology and Application
Online J Public Health Inform 2025;17:e72737
URL: https://ojphi.jmir.org/2025/1/e72737
doi:10.2196/72737

© Essence Lynn Wilson. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 5.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

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

XSL•FO
RenderX

Cool Solutions in Hot Times: The Case for Digital Health in Heatwave Action Plans

Maria Daniel Loureiro[1], MD; Neil Jennings[2], PhD; Emma Lawrance[2,3], PhD; Daniela Ferreira-Santos[1,4], PhD; Ana Luísa Neves[5,6], PhD

[1]Faculty of Medicine, Universidade do Porto, Porto, Portugal

[2]Grantham Institute - Climate Change and the Environment, Imperial College London, London, United Kingdom

[3]Climate Cares Centre, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[4]Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal

[5]CINTESIS@RISE, Centre for Health Technology and Services Research, Faculty of Medicine, University of Porto, Rua Dr. Plácido da Costa, Porto, Portugal

[6]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Ana Luísa Neves, PhD
CINTESIS@RISE, Centre for Health Technology and Services Research, Faculty of Medicine, University of Porto, Rua Dr. Plácido da Costa, Porto, Portugal

Abstract

This viewpoint highlights the critical need for proactive and strategic integration of digital health tools into heat-health action plans (HHAPs) across Europe. Drawing insights from the digital health surge during the COVID-19 pandemic and recent heat-related health impacts, we identify response gaps and suggest specific strategies to strengthen current plans. Key recommendations include leveraging mobile health communication, expanding telemedicine usage, adopting wearable health monitoring devices, and using advanced data analytics to improve responsiveness and equity. This perspective aims to guide policymakers, health authorities, and health care providers in systematically enhancing heat-health preparedness through digital health innovation.

(Online J Public Health Inform 2025;17:e66361) doi:10.2196/66361

KEYWORDS

climate change; digital health; telemedicine; disaster planning; health information systems

Introduction

Europe's three warmest years ever recorded have all occurred since 2020, with 2024 setting a new record as the warmest [1]. In 2024, Europe experienced an average temperature of 10.7°C, exceeding 1.5°C above the pre-industrial level—a threshold set by the Paris Agreement to significantly reduce the risks and impacts of climate change. Globally, heat stress reached unprecedented severity, with approximately 44% of the Earth's surface experiencing strong to 'extreme heat stress,' a significant increase over the historical maximum [2].

The escalating frequency and severity of heatwaves (periods of abnormally hot weather lasting from 2 days to months [3]) impose a major physiological strain on individuals, potentially resulting in acute illnesses such as heat exhaustion and heatstroke. Heatwaves also exacerbate the risks associated with non-communicable diseases such as respiratory diseases, mental health issues, cardiovascular and renal diseases, or diabetes [4,5]. The impacts of extreme heat are disproportionately felt by vulnerable populations, such as the elderly and individuals with pre-existing health conditions. Additionally, these severe temperature episodes substantially strain health care systems worldwide, increasing demand on emergency services, hospital admissions, and ambulance callouts [6-8].

Heat-health action plans (HHAPs) are structured, multisectoral frameworks [9] that outline preventive and responsive measures to reduce the health impacts of extreme heat events. In accordance with guidance from the WHO Regional Office for Europe, they should include aspects such as early warning systems, public information plans, health care preparedness measures, and real-time surveillance of heat-related health outcomes [10]. Their objective is to minimize heat-related morbidity and mortality, particularly among high-risk groups.

Despite the clear necessity and well-documented effectiveness of such plans, significant gaps remain in their implementation across Europe. According to recent assessments by the European Environment Agency, only about half—21 out of 38—of European countries currently have operational HHAPs [11]. Furthermore, significant variability exists in the scope, responsiveness, and implementation of existing HHAPs, with many plans lacking essential components such as real-time

XSL•FO
RenderX

health surveillance, targeted risk communication, or adequate integration with health care services. This heterogeneity has left large segments of the population insufficiently protected during extreme heat events. The limited effectiveness of these plans was evident during the summer of 2022, when, despite the activation of HHAPs across several European countries, more than 60,000 heat-related deaths were reported—a figure that underscores the partial and uneven success of current strategies [12]. Addressing these deficiencies by strengthening the adoption and quality of HHAPs is therefore not only urgent but ethically imperative to protect populations increasingly vulnerable to climate extremes.

To address this gap, this viewpoint aims to guide policymakers, health authorities, and health care providers in systematically enhancing heat-health preparedness through digital health innovation. Drawing from the digital health advancements that took place during the COVID-19 pandemic, we propose specific strategies for integrating targeted digital communication, telemedicine, and health data modeling into future HHAPs.

Digital Innovation in Heat-Health Preparedness: Insights From the COVID-19 Response

The COVID-19 pandemic highlighted a powerful example of how rapidly health care systems can innovate, particularly in terms of digital health tools [13]. The swift implementation of digital technologies across health systems, from mobile health apps and telemedicine to real-time data dashboards for disease monitoring, transformed public health management during the pandemic [14]. These tools were instrumental in improving patient surveillance, ensuring continuity of care, and supporting effective communication strategies between health authorities and the general population [15].

The pandemic experience also revealed key operational needs that digital interventions addressed effectively, which are also directly applicable to heat-health preparedness: the necessity of establishing direct channels of communication with the public, ensuring continuous access to health care services during crises, and developing the capacity to accurately model risks and adapt response strategies in real time.

To operationalize these lessons, a range of digital health strategies can be mobilized to strengthen existing plans and address current gaps in heatwave preparedness.

Rethinking Public Heat Communication Through Digital Channels

Effective communication is essential for raising awareness about the health impacts of heatwaves and promoting protective behaviors among the general public, vulnerable groups, caregivers, and organizations. Digital strategies can play a key role in sharing essential information such as identifying at-risk individuals (eg, those on psychoactive medications [16]), recognizing warning signs of heat-related illness, and locating available cooling centers. However, many countries continue to limit this information to institutional websites, which may not be easily accessible or sufficiently engaging to the broader population.

A more effective strategy would involve leveraging the digital infrastructure established during the COVID-19 pandemic to disseminate information [17,18]. National health authorities can repurpose mobile applications and social media channels to deliver critical information and issue public health guidance more directly and broadly. For example, during the pandemic, mobile applications were widely used to provide behavioral recommendations and assist with contact tracing [19]. Notably, the NHS COVID-19 app prevented approximately 1 million cases, 44,000 hospitalizations, and 9600 deaths during its first year of operation in England and Wales [20]. Similarly, social media platforms (such as Twitter) were used successfully by experts to disseminate public health messages at scale [21]. These tools could be readily adapted for heat-health communication. In addition, integrating messaging into health apps that are already in use by at-risk groups, such as for diabetes management or mental health platforms, could enable targeted outreach and ensure relevant information reaches those most in need [22,23]. Mobile health apps with monitoring functions, including those that track physiological data or geolocation, could be used to deliver personalized heat warnings based on user-specific risk profiles. Recent reviews have highlighted how such functionalities can enhance health communication and enable more proactive interventions during climate-related events [24]. The utilization of existing digital tools would not only leverage the substantial investments made in developing these apps but also enhance the responsiveness of health systems communication strategies to extreme weather events.

Telemedicine as a Climate Adaptation Tool

The use of telehealth services could play a vital role in maintaining essential care during heatwaves, a time when populations are advised to limit outdoor activities. By enabling remote consultations, telemedicine reduces exposure to heat during travel or in-person visits. Health professionals acknowledge that implementing telemedicine during heatwaves, particularly for patients with chronic conditions or reduced mobility, could significantly enhance clinical management through primary care [25].

In addition to improving access to care, virtual consultations have also been identified as a means to advance environmentally sustainable health care delivery [26]. Virtual care reduces unnecessary patient transport, limits physical infrastructure needs, and minimizes the carbon footprint of routine care activities. This dual benefit of improving both health outcomes and environmental sustainability strengthens the case for mainstreaming telemedicine in the context of climate-related health risks.

Despite its promise, telemedicine remains largely absent from most heat-health action plans in Europe, even in areas where telehealth infrastructure was rapidly developed during the COVID-19 pandemic [27]. This gap reflects a missed

XSL•FO
RenderX

opportunity to leverage proven tools in the face of increasing climate threats.

Moreover, telemedicine can play a critical role in pre-hospital triage, helping reduce pressure on overburdened emergency services during periods of high demand. For instance, during the COVID-19 pandemic, a structured multilevel telehealth service implemented in a resource-limited region in Brazil led to a 35% reduction in hospital admissions and a 40% decrease in mortality rates [28]. In contrast, during the 2011 heatwave, telecardiology support was successfully employed in Italy's Apulia region to evaluate patients with suspected acute heart disease, improving care while minimizing unnecessary hospital admissions [29]. Incorporating such strategies into HHAPs would improve health care system responsiveness during heatwaves and help ensure continuity of care for at-risk populations. Clear operational guidance on activating telehealth during extreme heat should be a core component of future planning.

Data Modeling for Adaptive Heat Responses

Heat action plans also should be capable of adapting to the dynamic interplay between health and heat. Integrating health data modeling and simulation into both the preparedness and response to phases can enhance the flexibility and precision of these plans [30]. To adequately capture the increased risk to vulnerable groups, activation thresholds for HAAPs should not only rely on ambient temperature, but also consider other meteorological parameters—such as humidity, nighttime temperatures, and heatwave duration—adjusted to the demographic and health profiles of the affected populations [31,32]. These may include factors such as age distribution, prevalence of chronic diseases, and proximity to health care services.

Combining health data with meteorological inputs also supports more informed policy making. This integration is essential for simulation and impact assessment studies, which can clarify the bidirectional relationship between causes and adaptations for extreme heat events. For instance, assessing the health benefits of air conditioning versus passive cooling methods in health care facilities requires balancing clinical effectiveness with environmental costs, energy use, and feasibility.

During heatwaves, near real-time indicators, such as emergency call volumes or hospital visits, paired with meteorological forecasts, can guide dynamic decision-making and allow for timely adjustments to response measures [33]. In France, the Heat-Health Watch Warning System is one example of such an integrated approach. It monitors both biometeorological indicators (which assess the impact of atmospheric conditions on living organisms) and health indicators such as mortality, emergency calls, and hospital visits for all causes and heat-related conditions. This comprehensive monitoring allows authorities, including civil protection and hospitals, to calibrate their response efforts [34]. Notably, during a severe heatwave in July 2006, the Heat-Health Watch Warning System saved more than 4300 lives in 18 days [33].

Integrating Digital Innovation Into Heat-Health Action Plans Is a Public Health Imperative

Despite the success of such systems, digital tools remain underutilized in most European HHAPs. Only a few countries incorporate mobile apps or digital platforms into their plans, mainly for information dissemination. In fact, most current strategies still rely on more traditional channels such as telephone hotlines (eg, in Germany, Belgium, Portugal, and Romania) or weather-warning websites (eg, in Slovenia and Italy) [35]. While institutional websites, radio, television, and, more recently, social media continue to play a role [9], they are limited in their ability to deliver tailored messages to high-risk groups or to support interactive, two-way communication. The lack of integration with digital health infrastructures leaves many HHAPs poorly equipped to respond to the demands of increasingly frequent and severe heat events.

To enhance health system preparedness for extreme heat events, there is an urgent need for forward-looking, adaptable strategies that harness the full potential of digital technologies. These not only enable faster and more targeted interventions but also support more anticipatory and scalable public health responses. Realizing this potential requires more than the deployment of new technologies; it demands a shift in public health strategy. A systems-thinking approach is essential, one that embeds digital innovation into the core of heat-health planning and ensures coordination across sectors and levels of care. Such an approach supports both rapid emergency response and long-term resilience in the face of escalating climate risks.

It is, however, imperative to critically assess the environmental impact of these technologies. While telemedicine and digital consultations can significantly reduce carbon emissions by minimizing patient travel [36], the broader digital infrastructure also contributes to greenhouse gas emissions. Data centers, which power digital health services, account for approximately 1% of global electricity consumption [37]. A comprehensive evaluation of the environmental footprint of digital health interventions is therefore essential.

To the same degree, the accessibility of digital health policies should be assessed. While these interventions may not reach everyone equally—particularly older adults or individuals with limited digital literacy [38]—they can free up capacity in the system. The broad implementation of digital solutions can improve accessibility by allowing more focused and personalized outreach to those who may benefit from direct or non-digital support. These tools may also reduce geographic inequalities by improving access to care for populations in remote or underserved areas. Interventions such as wearables for continuous vital signs monitoring or assisted teleconsultation do not require active technology use by the individual and can still be deployed effectively among these groups. Recent reviews have highlighted the growing role of wearable devices to monitor the health effects of heat events and support public health surveillance at scale [24].

XSL•FO
RenderX

Still, barriers such as language differences, limited access to digital devices, and unfamiliarity with technology remain significant challenges that may undermine the effectiveness of these interventions. Engaging community stakeholders in the design and deployment of digital health solutions can help ensure these interventions are culturally sensitive, user-friendly, and responsive to diverse population needs.

Integrating digital innovation into heat-health action plans is no longer optional; it is a public health necessity. As climate change accelerates the frequency and severity of extreme heat events, health systems must evolve accordingly. This viewpoint suggests that integrating digital tools into HHAPs is not simply an opportunity but an essential step toward modernizing public health preparedness in the face of climate change. We call on public health decision-makers to act swiftly and collaboratively to embed digital innovation within all stages of extreme heat response.

Acknowledgments

ALN is funded by the NIHR (National Institute for Health and Care Research) North West London Applied Research Collaboration and by the NIHR North West London Patient Safety Research Collaboration, with infrastructure support form the NIHR Imperial Biomedical Research Centre (NIHR203323). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Conflicts of Interest

None declared.

References

1. World Meteorological Organization. State of the global climate 2024. : World Meteorological Organization; 2025 URL: https://wmo.int/publication-series/state-of-global-climate-2024 [accessed 2025-08-01]
2. Copernicus Climate Change Service. Global climate highlights 2024. Copernicus. URL: https://climate.copernicus.eu/global-climate-highlights-2024 [accessed 2025-04-02]
3. Intergovernmental Panel on Climate Change. 'Glossary', in Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change: Cambridge University Press; 2023:2897-2930. [doi: 10.1017/9781009325844]
4. Ebi KL, Capon A, Berry P, et al. Hot weather and heat extremes: health risks. The Lancet 2021 Aug;398(10301):698-708. [doi: 10.1016/S0140-6736(21)01208-3]
5. Arsad FS, Hod R, Ahmad N, et al. The impact of heatwaves on mortality and morbidity and the associated vulnerability factors: a systematic review. Int J Environ Res Public Health 2022 Dec 6;19(23):16356. [doi: 10.3390/ijerph192316356] [Medline: 36498428]
6. Alho AM, Oliveira AP, Viegas S, Nogueira P. Effect of heatwaves on daily hospital admissions in Portugal, 2000-18: an observational study. Lancet Planet Health 2024 May;8(5):e318-e326. [doi: 10.1016/S2542-5196(24)00046-9] [Medline: 38729671]
7. Oberai M, Xu Z, Bach AJE, Phung D, Watzek JT, Rutherford S. Preparing for a hotter climate: a systematic review and meta-analysis of heatwaves and ambulance callouts in Australia. Aust N Z J Public Health 2024 Feb;48(1):100115. [doi: 10.1016/j.anzjph.2023.100115] [Medline: 38286717]
8. Thompson R, Lawrance EL, Roberts LF, et al. Ambient temperature and mental health: a systematic review and meta-analysis. Lancet Planet Health 2023 Jul;7(7):e580-e589. [doi: 10.1016/S2542-5196(23)00104-3] [Medline: 37437999]
9. WHO Regional Office for Europe. Heat and Health in the WHO European Region: Updated Evidence for Effective Prevention: WHO Regional Office for Europe; 2021. URL: https://www.who.int/europe/publications/i/item/9789289055406 [accessed 2025-08-01]
10. Matthies F, World Health Organization. Heat-Health Action Plans: Guidance: Euro Nonserial Publication; 2008. URL: https://iris.who.int/handle/10665/107888 [accessed 2025-08-01]
11. European Environment Agency. The Impacts of Heat on Health: Surveillance and Preparedness in Europe: Publications Office; 2024. [doi: 10.2800/8167140]
12. Romanello M, Napoli CD, Green C, et al. The 2023 report of the Lancet Countdown on health and climate change: the imperative for a health-centred response in a world facing irreversible harms. Lancet 2023 Dec 16;402(10419):2346-2394. [doi: 10.1016/S0140-6736(23)01859-7] [Medline: 37977174]
13. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 4;382(23):23. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
14. World Health Organization. Global Strategy on Digital Health 2020-2025, 1st edition: World Health Organization; 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2025-08-01]

XSL•FO
RenderX

15. Getachew E, Adebeta T, Muzazu SGY, et al. Digital health in the era of COVID-19: Reshaping the next generation of healthcare. Front Public Health 2023;11:942703. [doi: 10.3389/fpubh.2023.942703] [Medline: 36875401]
16. Martin-Latry K, Goumy MP, Latry P, et al. Psychotropic drugs use and risk of heat-related hospitalisation. Eur Psychiatry 2007 Sep;22(6):335-338. [doi: 10.1016/j.eurpsy.2007.03.007] [Medline: 17513091]
17. Kondylakis H, Katehakis DG, Kouroubali A, et al. COVID-19 mobile apps: a systematic review of the literature.. J Med Internet Res 2020 Dec 9;22(12):e23170. [doi: 10.2196/23170] [Medline: 33197234]
18. Sharma T, Bashir M. Use of apps in the COVID-19 response and the loss of privacy protection. Nat Med 2020 Aug;26(8):1165-1167. [doi: 10.1038/s41591-020-0928-y] [Medline: 32457443]
19. Khan HU, Ali Y, Azeem Akbar M, Khan F. A comprehensive survey on exploring and analyzing COVID-19 mobile apps: Meta and exploratory analysis. Heliyon 2024 Aug 15;10(15):e35137. [doi: 10.1016/j.heliyon.2024.e35137] [Medline: 39170132]
20. Kendall M, Tsallis D, Wymant C, et al. Epidemiological impacts of the NHS COVID-19 app in England and Wales throughout its first year. Nat Commun 2023 Feb 22;14(1):858. [doi: 10.1038/s41467-023-36495-z] [Medline: 36813770]
21. Rao A, Sabri N, Guo S, Raschid L, Lerman K. Public health messaging on Twitter during the COVID-19 pandemic: observational study. J Med Internet Res 2025 Feb 5;27:e63910. [doi: 10.2196/63910] [Medline: 39908546]
22. Paradis S, Roussel J, Bosson JL, Kern JB. Use of smartphone health apps among patients aged 18 to 69 years in primary care: population-based cross-sectional survey. JMIR Form Res 2022 Jun 16;6(6):34882. [doi: 10.2196/34882] [Medline: 35708744]
23. Guracho YD, Thomas SJ, Win KT. Smartphone application use patterns for mental health disorders: A systematic literature review and meta-analysis. Int J Med Inform 2023 Nov;179:105217. [doi: 10.1016/j.ijmedinf.2023.105217] [Medline: 37748330]
24. Koch M, Matzke I, Huhn S, et al. Wearables for measuring health effects of climate change-induced weather extremes: scoping review. JMIR Mhealth Uhealth 2022 Sep 9;10(9):e39532. [doi: 10.2196/39532] [Medline: 36083624]
25. Paganini M, Lamine H, Della Corte F, Hubloue I, Ragazzoni L, Barone-Adesi F. Factors causing emergency medical care overload during heatwaves: A Delphi study. PLoS ONE 2023;18(12):e0295128. [doi: 10.1371/journal.pone.0295128] [Medline: 38117826]
26. Pickard Strange M, Booth A, Akiki M, Wieringa S, Shaw SE. The role of virtual consulting in developing environmentally sustainable health care: systematic literature review. J Med Internet Res 2023 May 3;25:e44823. [doi: 10.2196/44823] [Medline: 37133914]
27. Gullslett MK, Ronchi E, Lundberg L, et al. Telehealth development in the WHO European region: Results from a quantitative survey and insights from Norway. Int J Med Inform 2024 Nov;191:105558. [doi: 10.1016/j.ijmedinf.2024.105558] [Medline: 39084085]
28. Oliveira CRA, Pires MC, Meira KC, et al. Effect of a structured multilevel telehealth service on hospital admissions and mortality during COVID-19 in a resource-limited region in Brazil: retrospective cohort study. J Med Internet Res 2024 Jun 10;26(1):e48464. [doi: 10.2196/48464] [Medline: 38857068]
29. Brunetti ND, Amoruso D, De Gennaro L, et al. Hot spot: impact of July 2011 heat wave in southern Italy (Apulia) on cardiovascular disease assessed by emergency medical service and telemedicine support. Telemed J E Health 2014 Mar;20(3):272-281. [doi: 10.1089/tmj.2013.0086] [Medline: 24404817]
30. Martinez GS, Linares C, Ayuso A, Kendrovski V, Boeckmann M, Diaz J. Heat-health action plans in Europe: Challenges ahead and how to tackle them. Environ Res 2019 Sep;176:108548. [doi: 10.1016/j.envres.2019.108548] [Medline: 31247429]
31. Yılmaz M, Kara Y, Çulpan HC, Can G, Toros H. Detection and regional analysis of heatwave characteristics in İstanbul. Sustainable Cities and Society 2023 Oct;97:104789. [doi: 10.1016/j.scs.2023.104789]
32. Dimitriadou L, Nastos P, Zerefos C. Defining heatwaves with respect to human biometeorology. the case of Attica Region, Greece. Atmosphere (Basel) 2021 Sep;12(9):1100. [doi: 10.3390/atmos12091100]
33. Pascal M, Laaidi K, Wagner V, et al. How to use near real-time health indicators to support decision-making during a heat wave: the example of the French heat wave warning system. PLoS Curr 2012 Jul 16;4:e4f83ebf72317d. [doi: 10.1371/4f83ebf72317d] [Medline: 23066514]
34. Laaidi K, Ung A, Wagner V, Beaudeau P, Pascal M. The French Heat and Health Watch Warning System: principles, fundamentals and assessment Environmental Health. 2013. URL: https://www.semanticscholar.org/paper/The-French-Heat-and-Health-Watch-Warning-System-%3A-%2C-Laaidi-Ung/74a4cbfe95b9abf74013efdf67c6d138898b2f0e [accessed 2024-06-11]
35. Casanueva A, Burgstall A, Kotlarski S, et al. Overview of existing heat-health warning systems in Europe. Int J Environ Res Public Health 2019 Jul 25;16(15):2657. [doi: 10.3390/ijerph16152657] [Medline: 31349585]
36. Purohit A, Smith J, Hibble A. Does telemedicine reduce the carbon footprint of healthcare? A systematic review. Future Healthc J 2021 Mar;8(1):e85-e91. [doi: 10.7861/fhj.2020-0080] [Medline: 33791483]
37. Jones N. How to stop data centres from gobbling up the world's electricity. Nature New Biol 2018 Sep;561(7722):163-166. [doi: 10.1038/d41586-018-06610-y] [Medline: 30209383]

XSL•FO
RenderX

38. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506. [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]

Abbreviations

HHAP: Heat-health action plan

Edited by E Mensah; submitted 15.09.24; peer-reviewed by J Campbell, Z Tomkins; revised version received 29.05.25; accepted 20.06.25; published 05.09.25.

Please cite as:
Loureiro MD, Jennings N, Lawrance E, Ferreira-Santos D, Neves AL
Cool Solutions in Hot Times: The Case for Digital Health in Heatwave Action Plans
Online J Public Health Inform 2025;17:e66361
URL: https://ojphi.jmir.org/2025/1/e66361
doi:10.2196/66361

© Maria Daniel Loureiro, Neil Jennings, Emma Lawrance, Daniela Ferreira-Santos, Ana Luísa Neves. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 5.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Remote Consultations in England During COVID-19: Challenges in Data Quality, Linkage, and Research Validity

Liliana Hidalgo-Padilla[1*], MSc; Massar Dabbous[1*], PhD; Kristoffer Halvorsrud[2], PhD; Thomas Beaney[3], PhD; Gideon Gideon[1], MSc; Eoin Gogarty[1], BEng; Geva Greenfield[3], PhD; Benedict Hayhoe[3], MD; Gabriele Kerr[3], MSc; Rosalind Raine[2], PhD; Nirandeep Rehill[2], MSc; Robert Stewart[1,4], FRCPsych; Fiona Gaughran[1,4], FRCPsych; Mariana Pinto da Costa[1,4], PhD

[1]Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, London, United Kingdom

[2]Department of Primary Care and Population Health, University College London, London, United Kingdom

[3]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[4]South London and Maudsley NHS Foundation Trust, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Mariana Pinto da Costa, PhD
Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, London, United Kingdom

Abstract

The COVID-19 pandemic accelerated the adoption of remote consultations across health care, requiring rapid adjustments in service delivery. Consequently, there is an urgent need to understand the impact of remote consultations on health pathways. This viewpoint paper explores key challenges in data sources in England that hinder research on the impact of remote consultations on health outcomes. Based on our experience conducting research on this topic, we present variations in observational study findings and their validity, considering differences in population characteristics and data sources. We provide recommendations to enhance data quality for future research, including improvements in data recording platforms and strengthened structures for linking primary and secondary care electronic health records.

(Online J Public Health Inform 2025;17:e66672) doi:10.2196/66672

KEYWORDS

telemedicine; telecare; remote consultations; data sources; telehealth; remote consultation; electronic health record; EHR; remote monitoring; England; data quality; secondary care; COVID-19

Introduction

The need to protect patients and health care staff during the COVID-19 pandemic accelerated the use of remote consultations in most health service settings [1-4]. This shift allowed medical appointments to be conducted remotely rather than in person, offering greater flexibility and convenience to both patients and providers [5]. Observational studies have examined the rate and impact of the shift to remote consultations during the COVID-19 pandemic across different populations and services, including mental health [6-9], oncology [10-14], dermatology [15,16], respiratory [17], orthopedics [18], older adult care [19], primary care [20], general outpatients [21], and COVID-19 inpatient care [22]. In the National Health Service (NHS) in England, a significant proportion of patient interactions continues to be conducted remotely despite the gradual reintroduction of face-to-face consultations.

Remote consultations have the potential to improve health care access, particularly for patients living in rural areas or those with limited mobility and constrained time availability, and to enhance efficiency amid increasing resource constraints. Although remote consultations are generally acceptable and convenient for both patients and health care professionals, there are disadvantages, such as the loss of nonverbal communication, inability to perform physical examinations, increased workloads for health care professionals, and insufficient institutional resources to support remote care effectively [3,23-26]. Understanding the long-term impact of remote consultations is crucial for developing strategies that optimize their benefits while addressing these limitations.

While evidence exists about the extent of ongoing health service provision for remote consultations and its impact on patient and health service outcomes, research in this area is hindered by inconsistencies in data recording, availability, quality, and reliability, as well as by variations in populations and methodologies, and difficulties linking primary and secondary care records.

The “Digital transformation in the NHS” strategy [27] was introduced to address these challenges by building knowledge from previous initiatives of technology implementation within

XSL•FO
RenderX

the NHS in England. This strategy emphasizes both the active engagement of staff and the critical need for up-to-date and reliable data to measure the impact of remote consultations on patient and health care outcomes. This will enable planning for improvements, address gaps, and expand or decommission service components, as appropriate, both across the NHS and in other health care settings.

In this viewpoint paper, we highlight challenges in the quality and linkage of electronic health record (EHR) infrastructures in NHS England, including inconsistencies in data documentation, interoperability issues, and limitations in data linkage between primary and secondary care. Additionally, we discuss variations in findings due to differences in population characteristics, service settings, and outcome measures. We also explore the uptake and engagement with remote consultations and the lack of long-term evaluations and consider the implications for future service planning. This paper stems from discussions among research teams from 3 leading universities in London, United Kingdom, who collaborated to explore the application of technology in the health care sector [28]. This viewpoint draws on our expertise working on data linkage across 2 geographical areas to evaluate the impact of remote consultations in different UK settings and on our systematic review of the relevant literature. While our focus is on England, similar issues may arise in other health care systems worldwide.

Data Challenges in Remote Consultations

Inconsistent and Incomplete Data Documentation

Complete, accurate, and consistent documentation is fundamental to ensure the quality and reliability of EHR-derived observational data. However, data recording inconsistencies pose a recurring challenge. Missing or variably recorded information can result from different factors such as time pressures during consultations, differences in clinician recording habits, and variations in health care settings [29]. For example, a study found that the frequency of diagnostic coding for long-term conditions was associated with patient sociodemographic characteristics, general practitioner (GP) practice—probably due to organizational processes—and disease inclusion in the Quality and Outcomes Framework financial incentive program [30]. However, it is hard to determine whether these associations are due to clinicians' biases or unmeasured confounding variables.

Additionally, large amounts of patient data are recorded as "free text" rather than structured data, making them difficult to analyze for research, service evaluation, and clinical audit. While natural language processing (NLP) techniques allow to extract valuable insights from free-text data [29,31], their effectiveness is limited by cross-system inconsistencies across EHR infrastructures and NHS trusts [32]. For example, NLP models need to be revalidated when used across other platforms due to different local documentation and recording practices. While the use of standardized and structured data (eg, predefined categories and drop-down boxes) is more amenable to analysis, it is insufficient because this does not guarantee improved accuracy or consistency between raters. Different platforms may offer different categorizations and therefore result in worse between-site consistency than NLP-processed clinical text [17].

Many of these challenges have been recognized as barriers in health care data research. However, the rapid shift to remote consultations during COVID-19 exacerbated these issues, with long-lasting effects [33]. Clinicians, navigating considerable operational pressures in the pandemic context, could have had less time to record properly structured data, leading to relying on free-text notes instead of structured coding.

EHR Interoperability Issues

The use of a plethora of different software and digital platforms for remote consultations with varying degrees of interoperability with EHRs also represents a challenge to deriving comparable data. Various software solutions used across NHS services often lack integration, preventing data exchange between primary, secondary, and community care settings. For example, discrepancies across different EHR providers could have led to inconsistent documentation of conditions such as long COVID, many of which were clinical- or service-related, making it difficult to derive meaningful comparisons. A study using population-based data in primary care from the 2 largest platforms in England (EMIS and TPP) [34] showed a higher number of people diagnosed with long COVID in London compared to the East of England. This regional variation could be associated with the differences in the user interfaces of both EHR providers. In addition, there was an increase in long COVID diagnoses in January 2021, probably due to the increase in awareness of this condition among clinicians and the inclusion of the diagnostic codes in at least 1 of the platforms.

Notably, the urgency of the COVID-19 pandemic led to the rapid adoption of new digital tools and platforms to facilitate remote consultations. While these systems addressed immediate needs, they often lacked full integration with existing EHR infrastructures, creating long-term challenges for data quality and interoperability.

Lack of Standardized Definitions for Remote Consultations Across Systems

Another significant barrier to evaluating remote consultations is the lack of standardization across platforms. Differing variable definitions can result in incomplete or incomparable data and pose a barrier to deriving meaningful findings. Definitions for what constitutes a remote consultation usually include telephone or videoconferencing interactions [6-8,18], but in some cases, it also considers email [17] or e-consultations [12], leading to discrepancies in data classification. Additionally, information is not always consistently recorded across systems, which may not allow a thorough understanding of events. For example, the inability to distinguish between different modes of psychiatric remote consultations (video, telephone, email, and SMS text message) hindered the possibility of exploring how clinicians performed clinical tasks based on the means of communication used [7], which might affect the quantity of information about a patient's "real-time" environment.

XSL•FO
RenderX

Difficulties in Data Linkage

Effective data linkage is essential to track patient journeys across primary and secondary care and assess the impact of remote consultations on their outcomes. However, in England, health care sectors use a wide range of EHR platforms, hence making linkage challenging [35,36]. The establishment of a network of Secure Data Environments aims to improve access to NHS data through shared and secure working practices, including sharing methods and coding [37]. This has been shown to be beneficial in studies linking primary and secondary care data to identify severe mental illness characteristics associated with cardiovascular diseases [38] but has been insufficient to compare attendance rates between remote and in-person consultations because modality was not available in primary care data [28].

Linkage of routine data resources can be complex, time-consuming, and expensive to establish. In part, this is because of the need for robust information governance. Without cross-system linkages of comparable data, it is difficult to assess long-term outcomes. For example, one study showed an increase in attendance rates to remote mental health services in South London, but the number of prescriptions of antipsychotics and mood stabilizers did not change [7]. However, one of the study limitations was that it did not use linkages to primary care data, which would be essential to have a more comprehensive understanding of the variations in prescription levels.

Data coverage and representativeness are limited when not all data can be linked. For example, a study examining primary care consultations for respiratory tract infections, using a large dataset covering 160 GP practices, only included patients with complete consultation notes and linked hospital records; thus, the authors warn that characteristics and outcomes of excluded patients may differ [17].

Consequences of Data Limitations on Research Quality

Variations in Study Findings

Similar research questions can sometimes yield conflicting results due to differences in data quality. This variability can be explained by differences in recording practices at the organizational (eg, funding and trust priorities), team (eg, staff levels and expertise), and individual patient and clinician levels (eg, number of sites, sample size, sociodemographic and clinical characteristics, and social disadvantage of the population). While the results are valid in each context, variability in the remits of nominally similar services all impact the interpretation and, thus, the generalizability of findings.

This can be illustrated by variations in the uptake of remote consultations among older people. Studies in mental health services in South London [7] and in primary care across England [19] during the pandemic demonstrated lower uptake of remote consultations among people aged 65 years and older compared to younger age groups. While this might suggest a general preference for face-to-face over remote among older patients or lack of access to technology, another study of patients with dementia in 2 mental health trusts reported fewer missed appointments in those attending remotely compared to face-to-face [8], which might be due to a more frequent presence of caregivers helping people with dementia in the use of technology.

Uptake could have also changed due to other patient characteristics across GP practices in England. Trajectories of 21 GP practices in Bristol, North Somerset, and South Gloucestershire between April-July 2019 and April-July 2020 varied by patient age (eg, no change for people aged 70 years and older), mental health status (increased among patients with poor baseline mental health and decreased among patients with good baseline mental health), and shielding status (increased in shielding patients and decreased in nonshielding patients) [20].

In terms of diagnostic accuracy, one study in primary care reported that remote consultations were sufficient to address most patient problems [20], but another study concluded that simplified digital consultation tool interactions with primary care patients could be only limited to low-risk queries [39]. The contrast in these findings might be secondary to differences in the study population (eg, patients from different regions of England) and in the data (eg, different data points and number of GP practices included). Furthermore, a study in secondary care settings analyzing oncology consultations reported that remote assessments were useful for routine follow-ups but insufficient for initial cancer diagnoses due to the lack of physical examinations [13].

Differences in outcomes may be attributable to different clinical and social needs when receiving care. For example, 3 studies explored remote triaging in 41 secondary care head and neck cancer services [14], 2 maternity services in hospital trusts [40], and 154 GP practices [41]. The triage in the head and neck cancer and maternity services, using data from May to August 2020 and October 2021 to February 2022, respectively, found that the remote triage was effective to identify the level of severity of risk cases, which was beneficial for the patients' care management, provided reassurance for staff, and alleviated infrastructural pressures. However, the study in primary care using prepandemic data did not find any differences in the assessment outcomes nor the time or burden to the practitioners. These differences might be explained by the impact of the context of the pandemic on both staff and patients, but also on the diverse needs for each type of service. Without comprehensive and standardized data, it is difficult to assess whether remote consultations lead to comparable patient outcomes as in-person visits.

Some studies suggest that remote consultations during the COVID-19 pandemic increased access to health care by reducing travel burdens [42-44] and improving attendance rates [45,46]. It was further suggested that insofar as the purpose was to exchange explicit and less emotionally loaded health care information, the use of telephone was particularly frequent, including check-up calls to patients [44,47]. Although these studies provide helpful discourses of facilitators and challenges during the implementation of remote consultations during the COVID-19 pandemic, these could be supplemented with quantitative data on the frequency of these consultation modalities.

XSL•FO
RenderX

Effects on Vulnerable Populations

Certain populations remain underrepresented in remote consultation research due to gaps in data collection and reporting. For example, the voices of patients with dementia were reportedly lost in remote communications, which primarily engaged their caregivers instead [47]. In studies with recently arrived migrants [48,49], patients with language barriers or living in limited-income settings found remote consultations hard to engage with [50].

Moreover, one study reported how the introduction of home-schooling increased the skills of using videoconferencing technology among affluent families from urban and rural settings, which facilitated access to remote health care [43]; however, poorer families often had to share computer facilities and operate with less stable broadband networks. Nevertheless, a study in Northwest London showed that appointment attendance rates were similar between sociodemographic groups, regardless of whether remote or in-person, indicating that remote consultations have neither widened nor reduced inequality [28]. However, these results might be biased, as only patients who had the ability and need to receive remote consultations were offered this modality. Without comprehensive data on these vulnerable populations, health care services risk implementing policies that fail to address their unique needs.

Gaps in Data Affecting Long-Term Evaluations

The lack of complete and standardized data hinders long-term evaluations of remote consultations. A key limitation of studies that investigated health care contacts by modality is that changes have primarily been studied early in the COVID-19 pandemic (up until September 2020), a time when many factors were affecting health care systems and uptake. Two studies explored changes in consultation rates over time in secondary mental health services [7] and primary care [20] between March-September and April-July 2020, respectively, with both reporting a decrease in face-to-face consultations.

Another study explored consultation rates and modalities in relation to socioeconomic deprivation from 2018 to 2022 [51]. The study used data from 397 GP practices in England and found that remote consultations increased from 0.91 on average per person-year in the period 2018‐2019 to 2.45 in 2020‐2021 and slightly decreased to 2.34 in 2021‐2022, suggesting changes in patient and clinician preferences. The study also showed that although remote consultations increased for all deprivation quintiles, it was larger for people in the most deprived quintile. However, inconsistencies in the recording consultation modality limited the ability to assess the reasons behind these changes, which could lead to an underestimation of the number of remote consultations. Similarly, a study analyzing NHS Digital data found that telephone consultations increased 3-fold between February 2020 and August 2021 [52]. People from low-income backgrounds were less likely to use remote consultations and receive same-day consultations at the beginning of the first COVID-19 lockdown in March 2020; however, these inequalities disappeared later in the pandemic period [52].

Public Perceptions of Remote Consultations

While data sources allow an understanding of the uptake of remote consultations, the public's perceptions could help put these findings in context. A content analysis of UK-based posts on the social media platform Twitter (subsequently rebranded X) [42] provided insights into patients' opinions regarding remote consultations during the different stages of the COVID-19 pandemic. The proportion of comments with a positive tone was almost double during the initial stages of the pandemic (March-May 2020) compared to both the period before the pandemic (January 2018-February 2020) and when some social distancing restrictions were lifted later in the pandemic (June-October 2020). Twitter was a useful data source to understand how public attitudes toward remote consultations evolve as a result of the different stages in the pandemic.

Patients reported that the choice of consultation modality (face-to-face, video, or telephone) is valued and varies according to the reason for the consultation (eg, a preference for obtaining repeat prescriptions over the phone) [43,53,54]. Patients and clinicians reported their discomfort with a lack of patient choice during the pandemic [46,53,55] and a concern that remote consultations might be used as a default option for cost- and time-saving reasons [20,49,50,55]. While the restrictions did not allow flexibility at the time, clinicians and patients can assess the advantages and disadvantages of each consultation modality in order to tailor each consultation for different patient needs.

Issues in detecting and evaluating variations in communication quality were present, including uncertainties of whether and how patients' queries were received and a lack of confidence using remote means of interaction [39,44]. Some patients reported that appropriate quality of care and relationships could only be built in person, in part due to the added value of nonverbal communication in face-to-face consultations [43-45]. Particular concerns were that remote consultations would increase responsibility on the part of patients to articulate their needs verbally (especially over telephone where patients cannot be seen) and reduce the availability of other informal or nonverbal avenues of communication provided in clinics and waiting rooms [39,42,44,45,49].

Recommendations and Conclusions

This paper, focusing on England, highlights the urgent need to improve data quality to better understand the variations in implementation, uptake, and impact of the wide-scale introduction of remote consultations as they become embedded. Many of the data challenges presented have relevance beyond the English context (eg, lack of reliability and standardization in EHRs and diversity in populations and services) [56-58]. Addressing issues related to data standardization, completeness, and linkage will enable more reliable evaluations and support evidence-based decision-making.

Missing, inconsistent, and inaccurate data in EHRs could be addressed by simplifying and automatizing platforms or software

XSL•FO
RenderX

to avoid adding an extra burden to clinicians when recording data. In addition, data sharing and standardization of variables and service definitions across platforms are a start to improving analytical processes and the reliability of outputs. Health care systems implementing remote consultations may find suggestions for improved data quality valuable while building or enhancing digital health care systems to ensure valid and reliable results from evaluations based on observational data. Furthermore, while linkage across platforms is a complex process, having standardized definitions across trusts could facilitate its outputs, which would positively impact the quality and representativeness of the data.

Qualitative evidence can help us understand barriers and facilitators to both patient and clinician affecting the use and uptake of remote consultations and may highlight data quality challenges (eg, poor quality of documentation and lack of information regarding choice). Due to safety reasons, remote consultations were the default option during the COVID-19 pandemic, but in a world that has become more familiar with remote interactions, understanding the patients' experience with this modality is essential to ensure a patient-centered approach. Comparative mixed methods studies of individuals and groups (including both patients and staff) experiencing different levels of social advantage according to a variety of demographic, socioeconomic, and geographical (eg, urban vs rural) parameters are also recommended to measure, quantify, and unpick the mechanisms by which remote consultation engagement and uptake changes over time across England.

From the later stage of the pandemic, blended or hybrid approaches incorporating remote options have emerged, so evaluations need to consider whether and when alternatives to remote contact are available in a patient's pathway. Decisions should also consider patient and service characteristics. Adding to complexity, but important for interpretation of outcomes, is service-level context on whether patients enter remote or nonremote pathways by their own choice or by that of clinicians or service planners, along with the factors used to guide such choices. Long-term and cost-effectiveness evaluations as well as the assessment of expected and unexpected benefits and adverse effects are needed to explore real-world impact and develop guidance for the future positioning of remote consultations.

A glossary of key terms is provided in Textbox 1.

Textbox 1. Glossary of key terms used in this article.

- Consultations: Appointments between a health professional and a patient to diagnose or treat a health condition.
- Data sources: The origin of the information gathered for research, including databases, studies, and health records.
- Digital transformation: The process of integrating digital technology into all areas of a health care system, changing how services are delivered, and how care is provided.
- Electronic health record: A digital version of a patient's paper chart, containing their medical history, treatment plans, and other health information.
- National Health Service: The publicly funded health care system in England that provides medical services to residents.
- Remote consultations: Medical appointments conducted via video, telephone, or digital platforms rather than in person.

Acknowledgments

The authors acknowledge input and support from the Remote Consultations Evaluation group. This report is independent research supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration (ARC) South London, the NIHR ARC North Thames, and the NIHR ARC Northwest London. LH-P is supported by a PhD studentship from the UK Research and Innovation (UKRI)-Economic and Social Research Council London Interdisciplinary Social Science Doctoral Training Partnership (LISS-DTP ES/P000703/1). NR was supported by NIHR ARC North Thames and the Beneficial Changes Network. KH is fully funded, and RR is partly funded by the NIHR ARC North Thames. BH, GG, and GK have been supported by the NIHR ARC North West London. GK was supported by the Beneficial Change Network. TB was supported by a clinical fellowship from the Wellcome Trust. FG and RS are supported by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London and the NIHR ARC South London at King's College Hospital NHS Foundation Trust. FG is also supported by the Medical Research Council (MRC)/UKRI Metabolic Psychiatry Hub. RS is also supported by the UKRI-MRC through the DATAMIND HDR UK Mental Health Data Hub (MRC reference: MR/W014386), the UK Prevention Research Partnership (Violence, Health and Society; MR-VO49879/1), an initiative funded by UK Research and Innovation Councils, the Department of Health and Social Care (England) and the UK devolved administrations, and leading health research charities. The views expressed in this publication are those of the authors and not necessarily those of the funders, including the National Institute for Health and Care Research or the Department of Health and Social Care.

Authors' Contributions

LH-P and MD wrote the original draft. MD and KH conducted the literature searches. MPC supervised this work. All authors reviewed and edited the manuscript.

XSL•FO
RenderX

Conflicts of Interest

BH declares that he also works for eConsult Health Ltd, a provider of electronic consultations for National Health Service (NHS) primary, secondary, and urgent or emergency care. In the past 3 years, FG has received honoraria for talks from Boehringer Ingelheim, Lundbeck, and Recordati. FG has a family member with previous professional links to Lilly and GSK. FG is in part supported by the National Institute for Health Research's (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, and the NIHR Applied Research Collaboration South London at King's College Hospital NHS Foundation Trust, and the Medical Research Council (MR/Z503563/1); and as part of Hub for Metabolic Psychiatry, one of the 6 hubs comprising the UK Research and Innovation Mental Health Platform (MR/Z000548/). Other authors have no conflicts of interest to disclose.

References

1. Khan N, Jones D, Grice A, et al. A brave new world: the new normal for general practice after the COVID-19 pandemic. BJGP Open 2020 Aug;4(3):32487520. [doi: 10.3399/bjgpopen20X101103] [Medline: 32487520]
2. Garfan S, Alamoodi AH, Zaidan BB, et al. Telehealth utilization during the Covid-19 pandemic: a systematic review. Comput Biol Med 2021 Nov;138(104878):104878. [doi: 10.1016/j.compbiomed.2021.104878] [Medline: 34592585]
3. Barnett P, Goulding L, Casetta C, et al. Implementation of telemental health services before COVID-19: rapid umbrella review of systematic reviews. J Med Internet Res 2021 Jul 20;23(7):e26492. [doi: 10.2196/26492] [Medline: 34061758]
4. Moynihan R, Sanders S, Michaleff ZA, et al. Impact of COVID-19 pandemic on utilisation of healthcare services: a systematic review. BMJ Open 2021 Mar 16;11(3):e045343. [doi: 10.1136/bmjopen-2020-045343] [Medline: 33727273]
5. Mittermeier S, Seidel A, Scheiner C, Kleindienst N, Romanos M, Buerger A. Emotional dysregulation and its pathways to suicidality in a community-based sample of adolescents. Child Adolesc Psychiatry Ment Health 2024 Jan 20;18(1):15. [doi: 10.1186/s13034-023-00699-4] [Medline: 38245793]
6. Chen S, Jones PB, Underwood BR, et al. The early impact of COVID-19 on mental health and community physical health services and their patients' mortality in Cambridgeshire and Peterborough, UK. J Psychiatr Res 2020 Dec;131(244-54):244-254. [doi: 10.1016/j.jpsychires.2020.09.020] [Medline: 33035957]
7. Patel R, Irving J, Brinn A, et al. Impact of the COVID-19 pandemic on remote mental healthcare and prescribing in psychiatry: an electronic health record study. BMJ Open 2021 Mar 30;11(3):e046365. [doi: 10.1136/bmjopen-2020-046365] [Medline: 33785494]
8. Hong JS, Sheriff R, Smith K, et al. Impact of COVID-19 on telepsychiatry at the service and individual patient level across two UK NHS mental health Trusts. Evid Based Ment Health 2021 Nov;24(4):161-166. [doi: 10.1136/ebmental-2021-300287] [Medline: 34583940]
9. Stewart R, Martin E, Broadbent M. Mental health service activity during COVID-19 lockdown: south london and maudsley data on working age community and home treatment team services and mortality from february to mid-may 2020. medRXiv. Psychiatry and Clinical PsychologyPreprint posted online on Jun 16, 2020. [doi: 10.1101/2020.06.13.20130419]
10. Cathcart P, Clayton G, Smith S, Dua S, Gandamihardja T. Virtual clinic triage of breast referrals during the COVID-19 pandemic: retrospective outcome analysis. Br J Surg 2022 Feb 1;109(2):e26-e28. [doi: 10.1093/bjs/znab424] [Medline: 34875021]
11. Kaddour H, Jama GM, Stagnell S, Kaddour S, Guner K, Kumar G. Remote triaging of urgent suspected head and neck cancer referrals: our experience during the first wave of the COVID-19 pandemic. Eur Arch Otorhinolaryngol 2022 Feb;279(2):1111-1115. [doi: 10.1007/s00405-021-07135-3] [Medline: 34661717]
12. Scott LJ, Murphy M, Price S, et al. Changes in presentations with features potentially indicating cancer in primary care during the COVID-19 pandemic: a retrospective cohort study. BMJ Open 2021 May 24;11(5):e050131. [doi: 10.1136/bmjopen-2021-050131] [Medline: 34031120]
13. Banerjee S, Voliotidis D, Parvin L, Rama SPK. Telephone triage of suspected head and neck cancer patients during the coronavirus disease 2019 pandemic using the Head and Neck Cancer Risk Calculator version 2. J Laryngol Otol 2021 Mar;135(3):241-245. [doi: 10.1017/S0022215121000657]
14. Hardman JC, Tikka T, Paleri V, ENT UK, BAHNO and INTEGRATE (The UK ENT Trainee Research Network). Remote triage incorporating symptom-based risk stratification for suspected head and neck cancer referrals: a prospective population-based study. Cancer 2021 Nov 15;127(22):4177-4189. [doi: 10.1002/cncr.33800] [Medline: 34411287]
15. Corden E, Rogers AK, Woo WA, Simmonds R, Mitchell CD. A targeted response to the COVID-19 pandemic: analysing effectiveness of remote consultations for triage and management of routine dermatology referrals. Clin Exp Dermatol 2020 Dec;45(8):1047-1050. [doi: 10.1111/ced.14289] [Medline: 32416014]
16. Patel NP. Remote consultations for patients with hidradenitis suppurativa during the COVID-19 pandemic: a single-centre experience. Clin Exp Dermatol 2021 Aug;46(6):1079-1081. [doi: 10.1111/ced.14687] [Medline: 33864282]
17. Dambha-Miller H, Hounkpatin HO, Morgan-Harrisskitt J, Stuart B, Fraser SDS, Roderick P. Primary care consultations for respiratory tract symptoms during the COVID-19 pandemic: a cohort study including 70,000 people in South West England. Fam Pract 2022 May 28;39(3):440-446. [doi: 10.1093/fampra/cmab127] [Medline: 34632504]

XSL•FO
RenderX

18. Shammout S, Wall R, Murphy PN, Jain K. Virtual clinics versus face-to-face review: is the benefit the same for new orthopaedic patients? Health Policy 2022 Jul;126(7):688-692. [doi: 10.1016/j.healthpol.2022.05.007] [Medline: 35644721]
19. Joy M, McGagh D, Jones N, et al. Reorganisation of primary care for older adults during COVID-19: a cross-sectional database study in the UK. Br J Gen Pract 2020 Aug;70(697):e540-e547. [doi: 10.3399/bjgp20X710933] [Medline: 32661009]
20. Murphy M, Scott LJ, Salisbury C, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177. [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
21. Tyler JMB, Pratt AC, Wooster J, Vasilakis C, Wood RM. The impact of increased outpatient telehealth during COVID‐19: retrospective analysis of patient survey and routine activity data from a major healthcare system in England. Health Plann Manage 2021 Jul;36(4):1338-1345. [doi: 10.1002/hpm.3185]
22. Francis NA, Stuart B, Knight M, et al. Predictors of clinical deterioration in patients with suspected COVID-19 managed in a "virtual hospital" setting: a cohort study. BMJ Open 2021 Mar 23;11(3):e045356. [doi: 10.1136/bmjopen-2020-045356] [Medline: 33757955]
23. Darley S, Coulson T, Peek N, et al. Understanding how the design and implementation of online consultations affect primary care quality: systematic review of evidence with recommendations for designers, providers, and researchers. J Med Internet Res 2022 Oct 24;24(10):e37436. [doi: 10.2196/37436] [Medline: 36279172]
24. Verma P, Kerrison R. Patients' and physicians' experiences with remote consultations in primary care during the COVID-19 pandemic: a multi-method rapid review of the literature. BJGP Open 2022 Jun;6(2):35031558. [doi: 10.3399/BJGPO.2021.0192] [Medline: 35031558]
25. Walthall H, Schutz S, Snowball J, Vagner R, Fernandez N, Bartram E. Patients' and clinicians' experiences of remote consultation? A narrative synthesis. J Adv Nurs 2022 Jul;78(7):1954-1967. [doi: 10.1111/jan.15230] [Medline: 35362191]
26. O'Hare D, Gaughran F, Stewart R, Pinto da Costa M. A cross-sectional investigation on remote working, loneliness, workplace isolation, well-being and perceived social support in healthcare workers. BJPsych Open 2024 Feb 26;10(2):e50. [doi: 10.1192/bjo.2024.7] [Medline: 38406834]
27. Health and Social Care Committee. Digital transformation in the NHS. : Parliamentary House of Commons; 2023 URL: https://publications.parliament.uk/pa/cm5803/cmselect/cmhealth/223/summary.html# [accessed 2025-08-01]
28. Kerr G, Greenfield G, Hayhoe B, et al. Attendance at remote versus in-person outpatient appointments in an NHS Trust. J Telemed Telecare 2025 Jun;31(5):721-731. [doi: 10.1177/1357633X231216501] [Medline: 38128925]
29. Shemtob L, Beaney T, Norton J, Majeed A. How can we improve the quality of data collected in general practice? BMJ 2023 Mar 15;380:e071950. [doi: 10.1136/bmj-2022-071950] [Medline: 36921932]
30. Beaney T, Clarke J, Salman D, et al. Identifying potential biases in code sequences in primary care electronic healthcare records: a retrospective cohort study of the determinants of code frequency. BMJ Open 2023 Sep 27;13(9):e072884. [doi: 10.1136/bmjopen-2023-072884] [Medline: 37758674]
31. Jackson RG, Patel R, Jayatilleke N, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ Open 2017 Jan 17;7(1):e012012. [doi: 10.1136/bmjopen-2016-012012] [Medline: 28096249]
32. Tayefi M, Ngo P, Chomutare T, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. WIREs Computational Stats 2021 Nov;13(6) [FREE Full text] [doi: 10.1002/wics.1549]
33. Lewis R, Smith J, Posaner R. Learning from the Pandemic Shift of Outpatient Services to a Remote Footing: A Rapid Evaluation Study: Birmingham Health Partners; 2024.
34. Walker AJ, MacKenna B, Inglesby P, et al. Clinical coding of long COVID in English primary care: a federated analysis of 58 million patient records in situ using OpenSAFELY. Br J Gen Pract 2021 Nov;71(712):e806-e814. [doi: 10.3399/BJGP.2021.0301] [Medline: 34340970]
35. Harron K, Dibben C, Boyd J, et al. Challenges in administrative data linkage for research. Big Data Soc 2017 Dec 5;4(2):2053951717745678. [doi: 10.1177/2053951717745678] [Medline: 30381794]
36. Harron K. Data linkage in medical research. BMJ Med 2022;1(1):e000087. [doi: 10.1136/bmjmed-2021-000087] [Medline: 36936588]
37. Goldacre B. Better, broader, safer: using health data for research and analysis: a review commissioned by the Secretary of State for Health and Social Care. : Department of Health and Social Care; 2022 URL: https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis [accessed 2025-08-01]
38. Woodhead C, Ashworth M, Broadbent M, et al. Cardiovascular disease treatment among patients with severe mental illness: a data linkage study between primary and secondary care. Br J Gen Pract 2016 Jun;66(647):e374-e381. [doi: 10.3399/bjgp16X685189] [Medline: 27114210]
39. Turner A, Morris R, Rakhra D, et al. Unintended consequences of online consultations: a qualitative study in UK primary care. Br J Gen Pract 2022 Feb;72(715):e128-e137. [doi: 10.3399/BJGP.2021.0426] [Medline: 34903520]
40. Bircher C, Wilkes M, Zahradka N, Wells E, Prosser-Snelling E. Remote care and triage of obstetric patients with COVID-19 in the community: operational considerations. BMC Pregnancy Childbirth 2022 Jul 8;22(1):550. [doi: 10.1186/s12884-022-04863-0] [Medline: 35804304]

XSL•FO
RenderX

41. Saunders CL, Gkousis E. Impact of telephone triage on access to primary care for people living with multiple long-term health conditions: rapid evaluation. Health Soc Care Deliv Res 2022;10(18):1-48. [doi: 10.3310/UCCE5549] [Medline: 35793423]
42. Ainley E, Witwicki C, Tallett A, Graham C. Using Twitter comments to understand people's experiences of UK health care during the COVID-19 pandemic: thematic and sentiment analysis. J Med Internet Res 2021 Oct 25;23(10):e31101. [doi: 10.2196/31101] [Medline: 34469327]
43. Greenhalgh T, Ladds E, Hughes G, et al. Why do GPs rarely do video consultations? qualitative study in UK general practice. Br J Gen Pract 2022 May;72(718):e351-e360. [doi: 10.3399/BJGP.2021.0658] [Medline: 35256385]
44. Paskins Z, Bullock L, Manning F, et al. Acceptability of, and preferences for, remote consulting during COVID-19 among older patients with two common long-term musculoskeletal conditions: findings from three qualitative studies and recommendations for practice. BMC Musculoskelet Disord 2022 Apr 2;23(1):312. [doi: 10.1186/s12891-022-05273-1] [Medline: 35366845]
45. Hughes L, Petrella A, Phillips N, Taylor RM. Virtual care and the impact of COVID-19 on nursing: a single centre evaluation. J Adv Nurs 2022 Feb;78(2):498-509. [doi: 10.1111/jan.15050] [Medline: 34590738]
46. Russell A, de Wildt G, Grut M, Greenfield S, Clarke J. What can general practice learn from primary care nurses' and healthcare assistants' experiences of the COVID-19 pandemic? A qualitative study. BMJ Open 2022 Mar 15;12(3):e055955. [doi: 10.1136/bmjopen-2021-055955] [Medline: 35292497]
47. Tuijt R, Rait G, Frost R, Wilcock J, Manthorpe J, Walters K. Remote primary care consultations for people living with dementia during the COVID-19 pandemic: experiences of people living with dementia and their carers. Br J Gen Pract 2021 Aug;71(709):e574-e582. [doi: 10.3399/BJGP.2020.1094] [Medline: 33630749]
48. Fu L, Lindenmeyer A, Phillimore J, Lessard-Phillips L. Vulnerable migrants' access to healthcare in the early stages of the COVID-19 pandemic in the UK. Public Health (Fairfax) 2022 Feb;203:36-42. [doi: 10.1016/j.puhe.2021.12.008] [Medline: 35026578]
49. Knights F, Carter J, Deal A, et al. Impact of COVID-19 on migrants' access to primary care and implications for vaccine roll-out: a national qualitative study. Br J Gen Pract 2021 Aug;71(709):e583-e595. [doi: 10.3399/BJGP.2021.0028] [Medline: 33875420]
50. Norman C, Wildman JM, Sowden S. COVID-19 at the deep end: a qualitative interview study of primary care staff working in the most deprived areas of England during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Aug 17;18(16):8689. [doi: 10.3390/ijerph18168689] [Medline: 34444437]
51. Vestesson EM, De Corte KLA, Crellin E, Ledger J, Bakhai M, Clarke GM. Consultation rate and mode by deprivation in English general practice from 2018 to 2022: population-based study. JMIR Public Health Surveill 2023 May 2;9:e44944. [doi: 10.2196/44944] [Medline: 37129943]
52. Green MA, McKee M, Katikireddi SV. Remote general practitioner consultations during COVID-19. Lancet Digit Health 2022 Jan;4(1):e7. [doi: 10.1016/S2589-7500(21)00279-X] [Medline: 34952678]
53. Antoun J, Brown DJ, Jones DJW, et al. Understanding the impact of initial COVID-19 restrictions on physical activity, wellbeing and quality of life in shielding adults with end-stage renal disease in the United Kingdom dialysing at home versus in-centre and their experiences with telemedicine. Int J Environ Res Public Health 2021 Mar 18;18(6):3144. [doi: 10.3390/ijerph18063144] [Medline: 33803708]
54. Pan-London research on the shift to remote consultations during the COVID-19 pandemic: lessons learnt. : NHSE&I London's Clinical Advisory Group; 2022 URL: https://www.arc-nt.nihr.ac.uk/media/lellw0vh/paper-4c-remote-consultations-report-for-lceg-240822-1.pdf [accessed 2025-08-01]
55. Sloan M, Lever E, Harwood R, et al. Telemedicine in rheumatology: a mixed methods study exploring acceptability, preferences and experiences among patients and clinicians. Rheumatology (Oxford) 2022 May 30;61(6):2262-2274. [doi: 10.1093/rheumatology/keab796] [Medline: 34698822]
56. Altman MR, Colorafi K, Daratha KB. The reliability of electronic health record data used for obstetrical research. Appl Clin Inform 2018 Jan;9(1):156-162. [doi: 10.1055/s-0038-1627475] [Medline: 29514352]
57. Brooks LA, Manias E, Bloomer MJ. A retrospective descriptive study of medical record documentation of how treatment limitations are communicated with family members of patients from culturally diverse backgrounds. Aust Crit Care 2024 May;37(3):475-482. [doi: 10.1016/j.aucc.2023.04.007] [Medline: 37339921]
58. Davis MDM, Schermuly A, Smith AKJ, Newman C. Diversity via datafication? Digital patient records and citizenship for sexuality and gender diverse people. Biosocieties 2023 Jun;18(2):451-472. [doi: 10.1057/s41292-022-00277-5]

Abbreviations

EHR: electronic health record
GP: general practitioner
NHS: National Health Service
NLP: natural language processing

XSL•FO
RenderX

Edited by E Mensah; submitted 20.09.24; peer-reviewed by G Koru, S Juddoo; revised version received 11.03.25; accepted 22.06.25; published 20.08.25.

Please cite as:
Hidalgo-Padilla L, Dabbous M, Halvorsrud K, Beaney T, Gideon G, Gogarty E, Greenfield G, Hayhoe B, Kerr G, Raine R, Rehill N, Stewart R, Gaughran F, Pinto da Costa M
Remote Consultations in England During COVID-19: Challenges in Data Quality, Linkage, and Research Validity
Online J Public Health Inform 2025;17:e66672
URL: https://ojphi.jmir.org/2025/1/e66672
doi:10.2196/66672

© Liliana Hidalgo-Padilla, Massar Dabbous, Kristoffer Halvorsrud, Thomas Beaney, Gideon Gideon, Eoin Gogarty, Geva Greenfield, Benedict Hayhoe, Gabriele Kerr, Rosalind Raine, Nirandeep Rehill, Robert Stewart, Fiona Gaughran, Mariana Pinto da Costa. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 20.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Technology for Mental Health: Reflections on Scope and Future Directions in Institutes of Higher Education in India

Seema Mehrotra[1], PhD; TK Srikanth[2], PhD; Neha Dahiya[3], MD; Pulkit Verma[4], MSc; Girish N Rao[5], MD; Prachi Sanghvi[6], PhD; Ashoo Grover[3], MD; Rajesh Sagar[7], MD

[1]Department of Clinical Psychology, National Institute of Mental Health and Neurosciences, Bengaluru, Karnataka, India

[2]E-Health Research Centre, International Institute of Information Technology, Bengaluru, Karnataka, India

[3]Division of Delivery/ Implementation Research, Indian Council of Medical Research (ICMR), New Delhi, India

[4]Informatics and Data Centre Division, Indian Council of Medical Research (ICMR), New Delhi, India

[5]Department of Epidemiology, Centre for Public Health, National Institute of Mental Health and Neurosciences, Bangaluru, Karnataka, India

[6]Department of Clinical Psychology, Rashtriya Raksha University, Gandhinagar, Gujarat, India

[7]Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, New Delhi, India

Corresponding Author:
Seema Mehrotra, PhD
Department of Clinical Psychology, National Institute of Mental Health and Neurosciences
MV Govindaswamy Centre
Bengaluru, Karnataka, 560029
India
Phone: 91 9448503853
Email: drmehrotra_seema@yahoo.com

Abstract

College years represent a pivotal phase as students transition into adulthood, a period marked by heightened vulnerability to mental health challenges. Beyond the high prevalence of common mental health issues, a large treatment gap (driven by both supply-side and demand-side factors) exacerbates the overall burden. Furthermore, students in higher education frequently experience psychological distress and subthreshold symptoms that impair well-being and daily functioning. Globally, technology-based mental health solutions have emerged as an important strategy to address unmet needs, with a growing evidence base across populations. Research has increasingly focused on examining digital mental health interventions for college students. Against this backdrop, we examine the challenges within India's large higher education system—which serves approximately 43 million students—and the expanding role of technology in this sector. We explore the potential for leveraging technology-based solutions to enhance student mental health initiatives within higher education institutions, considering relevant policies and guidelines that provide an impetus to these efforts. We reflect upon challenges and opportunities for implementing digital mental health interventions in Indian higher education, and propose strategic actions at institutional and governmental levels. Key considerations include data governance, safety, transparency, positioning of digital initiatives relative to in-person care, safeguards for content quality, provision of interventions at varying intensities, and recommendations for policy, governmental support, and research to optimize the use of technology for student mental health in institutes of higher education in India.

(Online J Public Health Inform 2025;17:e78065) doi:10.2196/78065

KEYWORDS

digital mental health; higher education; India; Institutes of Higher Education; student mental health; technology for mental health

Mental Health of Students in Universities

College years mark a crucial phase in the transition to emerging adulthood as individuals advance in their individuation journey, expand social connections, and explore their identities. These years also introduce numerous demands, including heightened academic pressures, career planning, adaptation to new social environments, formation of peer networks, and navigation of close relationships. Most college students experience stress in one or more life domains, such as finances, health, intimacy, relationships with family, relationships on campus, and problems experienced by significant others [1].

Youth and young adulthood constitute a vulnerable period, as 75% of mental health conditions have an onset before the age of 24 years [2]. The World Health Organization (WHO) World

XSL•FO
RenderX

Mental Health Surveys conducted in 21 countries with varied income levels indicated that 20.3% of college students had a 12-month prevalence of diagnosable mental health disorders [3]. Nearly one-third of first-year university students from 19 universities across 8 countries screened positive for at least 1 common mental disorder including anxiety, mood, or substance disorder as per the *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders* [Fourth Edition]*)* [4]. Beyond diagnosable conditions, subthreshold symptoms of common mental health conditions, such as depression, are also highly prevalent (eg, affecting up to 29% of youth) and can impair functioning, social activity, and performance, as well as increase the risk of major depressive disorder and suicide attempts [5]. An international study among undergraduate students across 12 countries in Europe, Asia, the Western Pacific, and North and Latin America found that 48% of participants exhibited clinically significant depressive symptoms [6]. In addition to common mental health conditions, rising youth suicide rates have drawn considerable attention, underscoring the urgent need for action. According to the WHO, suicide is the fourth leading cause of death among persons aged 15-29 years [7]. In India, the 18- to 29-year age group has the highest suicide deaths among both young men and women [8,9]. Student suicides constituted 7.6% of total suicides in the country, similar to the proportion among unemployed persons [8]. Over the last 10 and 20 years, total suicides increased by an average of 2% annually, whereas student suicides increased by 4%—twice the rate of total suicides [10]. For every death by suicide in India, more than 200 people experience suicidality in some form (eg, suicidal thoughts) and more than 15 suicide attempts occur [11]. A large-scale study across 9 Indian states found that one-fourth of college students had heard someone express suicidal ideas, 12.4% experienced suicidal thoughts in the past 12 months, and 6.7% reported a suicide attempt at some point in their lives; in the same study, one-third reported moderate-to-severe depressive symptoms and one-quarter reported moderate-to-severe anxiety [12].

There is voluminous research on the effectiveness of interventions for various mental health problems. For example, an umbrella review encompassing 31 meta-analyses and 608 primary studies evaluated psychological interventions for preventing and treating common mental disorders in college students and noted the availability of effective universal, indicated, and treatment-based interventions for both depression and anxiety [13]. Despite the availability of effective interventions, a substantial treatment gap persists globally and in India. Findings from the National Mental Health Survey of India 2015-2016 indicated a treatment gap of about 80% for common mental disorders [14]. A United Nations Children's Fund report (2021) highlighted that only 41% of youth aged 15-24 years surveyed in India felt the need to seek support when facing mental health challenges [15].

The World Mental Health Surveys further revealed that only 16.4% of students experiencing diagnosable levels of mental health problems received any treatment over a 12-month period [3]. In addition, a large international study of first-year students from 19 universities across 8 countries reported low treatment rates over the same timeframe: only 25.3% of students with mental health disorders and 29.5% of those experiencing suicidal thoughts and behaviors received care [16].

Not seeking help for common mental health issues in young people can lead to short- and long-term negative outcomes, including diminished educational and career prospects and an increased risk of developing comorbid conditions [17,18]. Multiple factors contribute to the treatment gap, including limited access to appropriate mental health care and insufficient or inequitably distributed resources (supply-side barriers), as well as demand-side barriers such as low perceived need, stigma, and adverse attitudes toward help-seeking [14,19]. Among students, additional barriers include stigma, fear of negative consequences for academic records, normalization of distress as a passing phase, preference for informal sources of help, low prioritization amid other commitments, skepticism about the utility of professional services, and inconvenient access [20,21].

These observations underscore the urgent need to address supply-side barriers by improving access to services as well as to tackle demand-side challenges by developing approaches that facilitate early identification and professional help-seeking. Past efforts have predominantly targeted access by reducing supply-side barriers, whereas demand-side obstacles have received comparatively less attention because of their complexity [22]. Given the multiple barriers to professional help-seeking documented among young adults, including students, there is a pressing need for innovative, culturally sensitive, acceptable, and scalable approaches to address unmet mental health needs [18,20,23].

Role of Technology for Mental Health in General and in Young Adults

Integration of technology into mental health services has expanded accessibility, outreach, and intervention delivery worldwide [24]. With increasing internet penetration and advances in digital tools—particularly in low- and middle-income countries (LMICs) such as India, there is an immense potential to address the mental health needs of diverse populations. Technology can also partly mitigate barriers related to shortages in the mental health workforce and enhance the scalability of interventions.

Globally, there has been an exponential increase in internet penetration, including in resource-limited countries. According to a joint report from the Internet and Mobile Association of India and KANTAR, the number of internet users in India has surpassed 800 million [25]. India is expected to achieve universal internet access by 2028-2029 [26]. Increased technology use can expand the reach and scope of mental health services [27]. Young adults, in particular, show strong interest in technology-enabled tools and online communication and often use them to obtain information and guidance about mental health concerns [28]. The internet also offers a platform for young adults to exercise autonomy and seek support for complex and sensitive issues [29].

Factors that influence adults' willingness to engage with electronic services include prior experience with online platforms, confidence in using technology, perceived need for

XSL•FO
RenderX

assistance, and the effectiveness and user-friendliness of apps [30]. Researchers have also examined students' perspectives on digital mental health interventions (DMHIs) in college settings, and findings suggest promising directions for implementation [31].

Technology-Based Interventions: Emergent Directions

A variety of technology-based interventions are being used and studied in the field of mental health. Examples include teletherapy; computerized, internet-based interventions; self-help mobile apps (including gamified tools); artificial intelligence (AI)–powered chatbots; virtual reality therapy; and online supportive communities. Solutions such as teletherapy have also been examined and there is some evidence to suggest that teletherapy achieves effectiveness comparable with in-person therapy for several conditions, including depression and anxiety [32].

Evidence supporting web-based and mobile app–based psychological interventions continues to grow for concerns such as stress, distress, depression and anxiety, eating disorder symptoms, and enhancement of well-being among students [33,34]. These programs have evaluated multiple approaches, including cognitive behavioral therapy (CBT), acceptance and commitment therapy, mindfulness-based techniques, and skills training delivered through digital platforms.

A systematic search of reviews and meta-analyses on DMHIs for university students found that web-based (computer-delivered) interventions were effective or at least partially effective in reducing depression, anxiety, stress, and eating disorder symptoms; similar findings were reported for mobile apps [35]. Another systematic review and meta-analysis of studies on the effectiveness of digital interventions for university students with ongoing mental health difficulties showed medium effect sizes for both depression and anxiety [36].

In addition to treatment, researchers have examined preventive and promotive interventions [37]. One review analyzed 52 publications from 15 countries on 48 mobile apps designed to promote positive mental health. It identified 39 randomized controlled trials and found some evidence supporting the effectiveness of individual apps for fostering well-being [38].

Technology-based platforms can also serve as a means to improve mental health literacy and reduce stigma. For example, while excessive social media use is associated with adverse mental health outcomes, studies also highlight its potential to promote mental well-being by raising awareness, building supportive networks, and offering early access to help for individuals in distress [39,40]. A qualitative study of youth described 3 features of social media consumption—social connection, engagement with content, and opportunities for self-expression—that may contribute to positive mental health [39]. A review of media-driven mental health awareness campaigns for individuals aged 10 to 24 years identified evaluations of 15 initiatives across 8 countries and reported generally positive changes in attitudes, beliefs, intentions (eg, reduced stigma), and behaviors (eg, increased help-seeking) [40].

Online anonymous screening programs are as yet another use of technology to reduce negative attitudes and stigma-related concerns among young adults [41]. Available research suggests that distressed young adults may prefer online modes for initial contact, which can increase subsequent willingness to pursue in-person services after initial online engagement [42]. Smartphones have also supported mental health education and peer support programs aimed at empowering communities and cultivating well-being [42].

There is also emerging literature on technology-enabled strategies to increase help-seeking among distressed individuals [43]. Targeted strategies to encourage help-seeking for mental health challenges appear effective for adults who are at risk or already experiencing difficulties [44].

Most research on DMHIs has originated in high-income settings. A review of 49 studies assessed mobile, online, and other remote technologies for treatment and prevention of mental disorders in LMICs and found that most studies focused on feasibility and acceptability and the results suggested that these hold promise for effectiveness. These studies encompassed various apps such as tools to enhance clinical care and train health workers, mobile solutions to identify mental disorders, digital platforms to improve treatment adherence and recovery, online self-help programs for individuals with mental health conditions, and programs for substance use prevention and management [45].

In India, research on technology for mental health has increased over the last decade. With the advent of the COVID-19 pandemic, state and central government initiatives expanded efforts to address mental health needs [46], and the field of telemedicine also advanced rapidly [47]. Studies from India show a gradual shift from screening-based electronic decision support systems, task-sharing models to increase contact coverage, and text message–based interventions to web-based and mobile app–based programs and integrated delivery modalities [48-55]. A WhatsApp-based intervention has shown effectiveness in reducing barriers and improving help-seeking inclination and behavior among distressed young adults, demonstrating promise as a scalable approach [56]. Technology-enabled training for the mental health workforce has also grown substantially [57].

The Government of India's Tele-Mental Health Program, Tele-MANAS (Tele Mental Health Assistance and Networking Across States) represents a national milestone in using technology to reduce the treatment burden and expand access to care [58]. However, additional research is needed to understand attitudes, preferences, and acceptability of DMHIs in low-income countries and among marginalized and underserved young people [59,60].

XSL•FO
RenderX

Technology for Mental Health in Higher Education Settings

Multiple barriers such as a preference for self-reliance, low perceived need, limited awareness of available services, understaffed on-campus student well-being services, scheduling constraints and conflicts, waitlists due to limited appointment availability, time constraints, low confidence in service utility, stigma, discomfort with professional visits, and confidentiality concerns can impede help-seeking even when services are available on campuses [61-63]. These observations have provided an impetus to the use of technology-based mental health interventions in universities worldwide.

Technology-based tools may include websites, mobile apps, and online platforms and forums to name a few. Some are unguided self-help resources (fully self-directed tools with no human support); while others incorporate human support that can range from minimal guidance by trained laypersons (via text, phone, or email) to blended formats that combine digital self-help with face-to-face or online clinician-delivered support (eg, therapy sessions). Unguided or minimally guided interventions may be helpful for mental health promotion and milder problems with less acute needs; self-guided DMHIs typically produce small effects [64,65]. More human resource–intensive, blended interventions may be appropriate for individuals with more severe symptoms or complex problems [55].

Young people frequently use the internet to seek information that helps them understand and address personal and health concerns [17,28]. This pattern has mixed implications. Despite frequent online searches about mental health, many students remain unaware of internet-based interventions [66]. In one study, students who searched online showed enhanced mental health literacy—greater optimism about psychotherapy and reduced stigma in some domains—but lower willingness to seek or recommend professional help [67]. A review of university students' perceptions and experiences with telemental health reported that students generally view it as convenient, accessible, user-friendly, and beneficial, with potential to mitigate stigma-related barriers [68]. Preferences for internet-based interventions with higher levels of human interaction, video content featuring humans have also been described [66].

Educational institutions can deploy technology-based platforms across the spectrum of prevention and care: universal interventions for all students, selective interventions for those at risk, and indicated interventions for students with significant symptoms. These efforts can support prevention, mental health promotion, treatment, and crisis response. Indicated interventions show the largest effects, particularly when participants have access to human support during the intervention, either in-person or via digital channels such as email or online contact [69].

A review of digital interventions for depression, anxiety, and psychological well-being among college students assessed effectiveness, usability, acceptability, uptake, and adoption [34]. Of 89 included studies, most interventions were website-delivered, with internet-based CBT being the most common approach. Many programs incorporated human support in the form of coaching; most demonstrated full or partial effectiveness on psychological outcomes, although nearly half did not report usability or acceptability. A separate systematic review evaluated the reach (percentage of invited students who expressed interest) and uptake (percentage of enrolled students who initiated an intervention) of CBT-based DMHIs for college students [21]. Of those studies that reported reach or uptake, most had evaluated unguided or guided self-help programs. Although the overall reach of DMHIs was found to be low as operationalized, enrolled participants demonstrated high uptake. Despite evidence that improving reach and uptake could increase public health impact, most studies did not report either outcome [21].

Another systematic review of 27 randomized controlled trials of technology-based interventions addressing depression, anxiety, and related concerns (excluding substance misuse and eating disorders) found that approximately half of the interventions yielded at least one significant positive outcome versus controls at postintervention. The review emphasized both the promise of such interventions in university settings and the need for more high-quality trials [70]. Digital technologies may also contribute to suicide prevention, although they are unlikely to be sufficient as stand-alone interventions and appear most effective when they complement ongoing clinical care [71].

It has been observed that DMHIs commonly used on campuses often lack robust, student-specific effectiveness data, underscoring the need for evaluations across diverse student populations and institutional contexts [72,73]. Additional research is needed to examine drivers and enablers of user uptake and engagement.

A comprehensive review examined 9 specific DMHIs offerings identified from a random sample of 200 colleges and universities in the United States through examining institutional website analyses and conducting interviews with administrators and experts from 20 institutions [73]. Interventions were grouped as multicomponent programs, self-guided programs, and connector programs. It was observed that states or colleges often purchase access for students to use these tools at no cost. Also, this review noted that some interventions were not designed specifically for college students but were widely used by them and frequently recommended by institutions. DMHIs have been implemented for screening, improving mental health literacy, and encouraging at-risk students to seek appropriate support. Other examples include online support groups, message boards, module-based web interventions, and skills-building apps focused on resilience, coping, and mindfulness.

Despite growing availability, a research-to-practice gap persists: technology-enabled interventions remain underused on many campuses, and few studies have evaluated their implementation within college counseling or wellness centers [74-77]. A systematic review of university students' digital use for mental health found that, although students are ready to use digital platforms for mental health and well-being, they face challenges related to discerning the trustworthiness of information and data safety [78]. In addition, the potential cost-effectiveness of

XSL•FO
RenderX

DMHIs in college and university settings remains an area that is yet to be examined systematically [73].

Indian Education System: Challenges and Opportunities in the Use of Technology for Mental Health

India has one of the largest education systems in the world, with 1168 universities and university-level institutes, 45,473 colleges, and 12,002 stand-alone institutes, including 165 institutes of national importance. On average, there are 30 colleges per lakh (100,000) eligible persons aged 18-23 years in India [79]. In 2021-2022, 78% of colleges in India were found to be privately operated and 66% of total college enrollment occurred in these institutions, indicating higher education costs for students [80]. Total enrollment in higher education was nearly 43.3 million in 2021-2022, according to the All India Survey on Higher Education [79]. The National Education Policy (NEP) aims to increase the gross enrollment ratio to 50% by 2035, which would place a substantially larger proportion of the eligible population in the higher education system and has implications for ensuring accessible mental health support across institutions [81].

India has experienced rapid digitalization across many domains, including higher education (eg, course orientations, electronic learning, content sharing, evaluation and feedback systems, and online guidance). As one of the world's most populous nations with a substantial youth population, India faces unique opportunities and challenges in education and health, including providing quality higher education and health access to a large and diverse student body. The higher education sector is increasingly integrating technology to support remote learning and digital classrooms. The expansion of internet connectivity and mobile technology has accelerated the rise of online learning platforms, virtual classrooms, and Massive Open Online Courses, allowing students to access materials and participate in interactive learning [82,83].

Despite notable progress in strengthening higher education in India, several challenges exist that have a bearing on student mental health and well-being. Two interrelated issues are employment opportunities and an exam-centric system. According to the International Labour Organization, more than half of employed graduates in India work in low-skilled jobs, and fewer than one-third hold positions requiring high-level skills; many highly educated young adults, including those with technical training, are overqualified for their jobs [84].

India's examination-centric education system has been widely criticized for prioritizing rote learning over competency and skill development, influencing career paths and contributing substantially to student stress. The National Eligibility cum Entrance Test and the Joint Entrance Examination are among the most competitive entrance examinations, viewed as critical steps toward admission to medical and engineering programs. Millions of students compete for a limited number of seats in top institutions, fueling intense pressure and stress. This dynamic has also driven the growth of a coaching industry focused on examination preparation, with students sometimes investing years in study and attending coaching institutes for extended periods [85].

These pressures carry financial, social, and health consequences. Rigorous preparation schedules and frequent mock examinations often reduce time for social interaction, physical activity, and self-care, while increasing risk for anxiety and depressive symptoms because of self-doubt, overwhelming stress, fear of failure, setbacks, uncertainty about career progress, guilt related to perceived underperformance, concerns about disappointing or burdening families, and a low sense of support [85,86]. These trends—combined with employability challenges, societal norms that prize a narrow range of careers, the high prevalence of common mental health concerns (including subthreshold symptoms) among young adults, a shortage of mental health professionals, and low awareness and stigma around help-seeking—further exacerbate the mental health situation in Indian higher education.

Potential for Using Technology to Support Mental Health in Indian Institutes of Higher Education and Key Considerations: Policy Support and Opportunities for Integration

The University Grants Commission of India has issued Guidelines for the Promotion of Physical Fitness, Sports, Students' Health, Welfare, and Psychological and Emotional Well-being [87]. These guidelines recommend that every institute of higher education establish a Student Services Centre to provide counseling, guidance, and mental health services in both online and offline formats, including telehelp [87].

India's NEP aims to enhance student learning outcomes, improve teacher training, and strengthen educational infrastructure [81]. From 2013-2014 to 2020-2021, combined state and central government education expenditure ranged from 3.9% to 4.6% of gross domestic product; the NEP recommends raising education spending to at least 6% of gross domestic product [88]. Several NEP recommendations could function as protective factors for student well-being, including a shift toward continuous assessment rather than exclusive emphasis on end-of-course examinations and multiple entry and exit options in undergraduate programs that allow students to realign academic trajectories based on interest and opportunity [88]. The NEP's emphasis on digital tools and online learning platforms creates opportunities to not only enhance educational access and academic performance but also to expand access to mental health support through technology-enabled interventions.

The NEP's governance vision—independent institutional boards composed of highly qualified members, greater academic and administrative autonomy, and a "light but tight" regulatory framework under a single higher education regulator—may help embed mental health initiatives into institutional offerings. However, there are likely to be implementation challenges including resource constraints, acceptability and receptivity among stakeholders, and administrative bottlenecks.

XSL•FO
RenderX

India's National Suicide Prevention Strategy, released in 2022, is another policy document of high relevance. It recognizes suicide as a major public health concern and sets a goal of reducing suicide mortality by 10% by 2030 [89]. The strategy calls for a multisectoral approach and highlights the need for multipronged approaches including efforts to raise awareness of common mental health concerns, reduce stigma, strengthen psychosocial support, train key stakeholders (eg, gatekeeper training of teachers) for early risk identification and referral, expand the mental health workforce, improve help-seeking among persons in distress, and bolster supports for individuals experiencing suicidal crises and their families and using youth engagement programs to name a few. [89]. Many of these recommended activities could benefit from technology-based platforms for scalable implementation.

The recent Digital Personal Data Protection (DPDP) Act has direct implications for mental health data privacy in India [90]. It seeks to protect personally identifiable mental health information from misuse, discrimination, and exploitation and mandates that the data fiduciaries (the entities processing the data) must obtain informed consent from individuals and should process data for specified purposes and adhere to principles such as data minimization and retention. Moreover, individuals have the right to withdraw consent, which obligates data fiduciaries to stop collecting, storing, using, or sharing data unless permitted under the Act or other relevant applicable laws [90]. Complexities in operationalizing these provisions (particularly within [Institutes of Higher Education] IHEs) require careful consideration. In parallel, the Indian Council of Medical Research has issued Ethical Guidelines for Application of AI in Biomedical Research and Health Care, which outline principles for ethical development, deployment, and governance of AI-based solutions in research and clinical practice, emphasizing accountability, autonomy, data privacy, risk minimization, and fairness [91]. While there is no comprehensive regulatory body covering mental health apps, a number of legal and regulatory frameworks would be relevant depending on the nature of the digital intervention. Going forward, it would be useful to evolve a uniform policy on the implementation of technology-based platforms for mental health in educational settings based on applicable legal and regulatory frameworks.

Implementation Challenges and Enabling Measures for DMHIs in Higher Education

Both digital and in-person interventions for college students may pose somewhat distinct barriers to their usage. While barriers such as procedural hurdles in scheduling, including time constraints, financing, lack of trust in confidentiality, fear of stigma or social consequences, become salient for in-person interventions, content-related dissatisfaction, unmet needs for direct human communication, and privacy concerns are some of the key deterrents to digital interventions experienced by students [92].

A review of research from 2019 to 2024 on DMHIs for college students revealed that despite promising psychological outcomes and good adherence, there was only modest uptake, with most studies originating from the United States or Canada and featuring predominantly female samples [93]. A scoping review of DMHIs for adolescents in LMICs found that while fear of exposure, stigma, and low awareness remained significant barriers, factors such as free access, confidentiality assurances, culturally appropriate content, and inclusion of human-led options facilitated DMHI engagement [94]. A Delphi study to attain expert consensus on barriers and strategies to improve uptake of DMHI for college students from diverse and marginalized backgrounds in the United States indicated strategies such as co-design processes, user-customizable privacy settings, content personalization, simplified onboarding, and embedding DMHIs within campus ecosystems. The study also highlighted the value of emphasizing autonomy, flexibility, and inclusion of representatives from the target groups to enhance acceptability and uptake of digital interventions [95].

There is a notable dearth of research in LMICs on the development and implementation of digital interventions, with an even greater gap in the context of IHEs. While there remains an urgent need for hybrid effectiveness–implementation studies to address the significant research gap in LMICs [93], existing international evidence on barriers and facilitators can help guide context-sensitive strategies for Indian higher education settings. The acceptability and uptake of digital interventions, as well as engagement with them, may be enhanced by measures such as student involvement across development, deployment and evaluation phases, transparent information on the scope of digital interventions and privacy policies, and customizable privacy controls that signal safety and credibility [95]. Offline-first apps which can perform core functions without internet access, may help minimize barriers such as bandwidth limitations and intermittent connectivity [96]. Emphasizing autonomy, flexibility, and representation of target groups in campaigns that build awareness about digital and in-person interventions may further increase their appeal among diverse groups, including marginalized students and reduce barriers to help-seeking [95]. A simplified onboarding process, clear in-app explanations of features and risks, and human support (such as brief coaching or scheduled check-ins) can mitigate barriers related to digital literacy for college students. Scope for content personalization, adoption of stepped-care models that offer varying levels of intervention intensity based on individual needs, and integration of DMHIs within campus ecosystems—such as counselling centers, and peer support networks can enhance student engagement with mental health services [73,97]. These efforts are more likely to succeed when paired with clear, accessible messaging about available support and referral pathways, helping students navigate both digital and in-person options with greater ease [20,93].

Linguistic and sociocultural diversity is another challenge in implementing DMHIs in Indian IHEs. Even when mental health content is translated, complex terminology can impede understanding. A few practical language-access strategies can improve comprehension and engagement. Approaches such as using locally meaningful terms, culturally appropriate adaptations, and offering human support to help students navigate app content can be helpful. Given the prominence of English in higher education alongside regional languages, and

XSL•FO
RenderX

frequent code-switching in everyday student communication, mixed-language content tailored to context may be more effective than strictly monolingual formats in DMHIs. Incorporating audio content in local languages, delivered in a conversational style, may further enhance accessibility, and this warrants further research.

The rapid proliferation of mental health apps, often developed with minimal involvement from qualified mental health professionals [98], raises concerns about the choice of digital tools to be deployed in IHEs, keeping in view their validity, safety, and user trust. Ethical issues related to potential harm, inadequate safeguards, and unclear accountability can further complicate implementation [99-101]. Additionally, the financial burden of sustaining these interventions at scale, especially in resource-constrained settings, presents a significant barrier. These issues are discussed in greater detail in the subsequent section on strategic government-level action.

Reflections on Strategic Actions Across Levels: Key Considerations for IHEs

Data Governance and Terms of Use

IHEs should establish clear data governance policies and accountability mechanisms to ensure a safe digital environment for students seeking digital mental health services. Students must receive transparent, easy-to-understand information about ownership and use of personally identifying information and sensitive mental health data. Institutions should implement specific, clearly articulated terms and conditions and consent protocols [20]. These should address, at a minimum, conditions for any data sharing with the institution; the scope and limits of the intervention; and the use of AI and its limitations. To enhance safety on online mental health platforms, institutions should adopt appropriate data-handling policies, establish risk-management protocols for monitoring signs of distress, and proactively offer support when needed.

Positioning of DMHIs and Integration with In-Person Services

Current evidence on DMHIs in IHEs, notable gaps in mental health and digital literacy as well as expert consensus, suggests that these interventions are best positioned as additions or supplements to face-to-face professional services rather than as stand-alone substitutes [73,102-104]. Framing DMHIs in this complementary role can also enable the development of stepped-care models, where students can access interventions of increasing intensity based on their needs and blended approaches that combine digital tools with in-person professional support, both of which are emerging as promising strategies within IHEs [20,55]. Accordingly, digital offerings should incorporate features such as timely, algorithm-driven nudges to encourage professional care when indicated [98]; clear crisis-support guidance and multiple help-seeking options; integration with offline services; and transparent communication about the platform's capabilities and limitations.

Ensuring Content Quality of DMHIs

The rapid proliferation of mental health platforms and mobile apps—varying widely in quality, purpose, and clinical grounding—can overwhelm and confuse users, while rigorously evaluated tools often remain out of reach [60,98]. Until stronger evidence accumulates, IHEs could prioritize platforms that are developed or vetted by qualified mental health professionals and include clinical oversight to the extent feasible. This approach can help ensure delivery of evidence-informed content, minimize potential harm, and maximize therapeutic benefit for the users.

Collaborative Engagement With Students Across Program Phases

Active student involvement in the development, decision-making, and implementation of mental health programs and services—both digital and offline is essential for their success [17,20,73]. Service offerings perceived as tokenistic, externally imposed, or disconnected from an ethos of shared responsibility for student well-being are unlikely to be adopted. To support meaningful collaboration, IHEs should engage students across all phases—from co-design to rollout and evaluation and incorporate regular feedback mechanisms, such as periodic surveys, to refine services and digital tools, from users as well as nonusers of services.

Provisioning for a Range of Apps in IHEs

Institutions should plan for training platforms and resources for both educators and students, as outlined in the sections below.

Use of Online Platforms for Regular Orientation and Training for Leaders and Faculty

Leadership development, facilitation, and sensitization—together with training that equips faculty to support socioculturally diverse student groups—are important components of student mental health programs [105,106]. Faculty training for student mental health and suicide prevention may include mentoring vulnerable students; recognizing suicide risk and warning signs; and establishing processes for early identification, first-line support, crisis intervention, and appropriate referral. Although technology platforms in Indian higher education currently focus on teaching, learning, and evaluation—as well as on planning and administrative processes—these platforms can be adapted to deliver structured modules that help educators and leaders understand student mental health needs, provide tools for support, and guide them in mentoring, risk assessment, and offering assistance for students in distress.

Spectrum of Technology-Based Resources for Student Mental Health

IHEs can deploy a diverse range of technology-based mental health initiatives that offer students a sense of choice and autonomy and allow them to engage with supports aligned to their felt needs, preferences, and readiness for help. These may include (1) centralized digital repositories of mental health resources in multiple formats and languages, regularly updated and publicized. (2) Interactive platforms to build mental health literacy, reduce stigma, and foster inclusion and belonging. (3)

XSL•FO
RenderX

Anonymized self-screening tools with tailored recommendations on intervention intensity and access pathways. (4) Peer support training tools and moderated online forums to strengthen student-led support systems. (5) Low-intensity chat-based support, including AI-assisted options, moderated by trained nonspecialists. (6) Minimally guided self-help tools for common concerns and personal well-being goals. (7) Targeted digital outreach (eg, via WhatsApp) to reduce help-seeking ambivalence among distressed, non–treatment-seeking students). (8) Blended care models combining digital self-help with online or offline therapist sessions for students needing higher-intensity support. (9) Access to telehelplines offering crisis support and brief interventions.

Beyond access to different kinds and intensities of interventions, offering format choices (eg, text- vs audio-based content) and options for communicating with a trained human agent (eg, chat- vs phone-based support) can increase the appeal of DMHIs for college students. The approaches listed above need to be integrated within campus ecosystems and aligned with broader mental health strategies rather than operate in isolation. Whenever feasible, these could be loosely integrated or "bundled" to allow students to move from one to another and exercise rights to indicate preferences and consent for transferring personal data between apps to improve efficiency. Students should be empowered to exercise their rights by explicitly indicating preferences and providing informed consent for any transfer of personal data between apps or services. This can facilitate both efficiency and respect for autonomy, including the right to opt in or out of such data-sharing arrangements. In addition to user-initiated transitions across levels, DMHIs need to incorporate algorithmic nudges that prompt students to consider interventions that are more in alignment with the nature and severity of their concerns. This function of DMHIs becomes critical in the context of low mental health literacy.

Indicators of Success

The implementation knowledge base benefits from systematic use of metrics that capture delivery realities alongside clinical outcomes and provide leads for further planning. Campuses can track practical indicators when embedding DMHIs within institutional support systems. Reach, adoption, effectiveness, implementation, and maintenance (often referred to collectively as RE-AIM) are widely used parameters for evaluating DMHIs for college students [93]. When interventions are integrated into campus mental health services, indicators of success may include increased overall uptake of services; the proportion of previously distressed, non–treatment-seeking students who initiate low-intensity digital supports (particularly those designed to enhance help-seeking); sustained engagement over time; user satisfaction; feasibility and acceptability among stakeholders; and evidence of sustainability within the campus context, in addition to relevant effectiveness outcomes.

Key Considerations for Government Action

1. Developing a coherent set of national guidelines for mental health apps—both broadly and for use within higher education can incentivize local developers to work toward creating quality solutions. For IHEs, such guidelines should specify minimum standards for content quality, mechanisms for independent quality checks, and safeguards for user safety. Clear protocols for data governance, privacy, and security and regulatory oversight to ensure ethical deployment in IHEs would also be essential.
2. In a country like India, with a vast population of college-going young adults and a rapidly growing number of newer digital apps and services entering the marketplace, it can become very confusing and overwhelming for inclined users to find an app that would suit their needs [98]. A centrally managed navigator platform that curates evaluated options, publishes plain-language summaries, and helps students compare and select appropriate digital interventions can ease the burden on the users to find what may suit their needs. Beyond navigation, such a comprehensive, youth-focused digital platform across IHEs could also serve as an essential resource to provide credible self-help materials in multiple languages, and feature a directory of affordable mental health services for students. Additionally, such a platform could include positive peer experiences of help-seeking and thus play a role in shaping peer norms around help-seeking for mental health.
3. Scarcity of trained mental health professionals in India, coupled with observations on low uptake of counselling services in IHEs and students' preference for online mediums, highlights the utility of blended offerings that combine offline services with digital tools. Training of mental health service providers in IHEs in the effective and ethical deployment and integration of DMHIs should be prioritized as part of a coordinated government-led capacity-building initiative. This effort can itself leverage online training modules to ensure scalability and reach.
4. Recognizing and incentivizing IHEs through accreditation, or awards based on well-functioning student mental health support systems, can encourage greater investment in mental health initiatives including DMHIs, and lead to the development of best practices and improved adoption.
5. To operationalize the above-mentioned directions at a national scale, structured multistakeholder collaborations are warranted. Lack of coordinated involvement of key actors (eg, academics, health providers, end users, and private industry) has been identified as a barrier to implementing digital mental health solutions [107,108]. India's digital mental health landscape similarly reflects fragmented growth, with isolated efforts from academia, government, private industry, and individual innovators [109]. Consolidating these efforts through formal collaboration can address fragmentation and accelerate the responsible scale-up of DMHIs for IHEs in India.
6. Reviews of DMHIs in IHEs have highlighted a consistent gap in reporting on implementation adaptations and

XSL•FO
RenderX

cost-related data—even in studies conducted in high-income settings [73]. DMHIs vary in cost and cost-effectiveness depending on their nature and context of implementation [109,110]. Moving forward, it is critical to examine how DMHIs can be strategically integrated with in-person services to optimize cost-efficiency and enhance the overall effectiveness of mental health support systems. Mandates by the government to report costs, uptake and engagement, and satisfaction with mental health services by IHEs can improve this scenario for the government to evolve strategies for optimizing resources while expanding access to effective mental health support for students.

7. Sustainable financing is critical to expanding access to student support systems within IHEs. This could include increased public investment in mental health, establishment of dedicated budget lines for student-mental health services, and the exploration of blended-finance models that combine government, philanthropic, and institutional resources to ensure long-term viability and reach. Public-Private-Partnership (PPP) models are recognized facilitators for integrating digital technologies within mental health systems [111]. In India, relevant precedents include PPPs in the National Dialysis Programme, Gujarat's Chiranjeevi Yojana, and telemedicine initiatives such as Apollo Telehealth in collaboration with state governments [112-114]. Applied to DMHIs for higher education, PPPs can balance public interest with commercial viability, enable evidence-informed planning, and support sustainable implementation costs through shared resourcing. This would require a national liaison or coordination agency (or an existing national body assigned this role) which could serve as a single point of contact among stakeholders.
8. Amid rising concerns around student mental health in India, underscored by the recent Supreme Court of India directives for IHEs regarding access to mental health services [115], there is an urgent need for dedicated funding to support national-level implementation–effectiveness research on integrated mental health strategies, including DMHIs. This effort should be embedded within a long-term operational framework, rather than a time-bound project, to ensure sustained progress in research, policy development, and system-wide integration.

On the whole, a multitiered framework of technology-based, evidence-informed mental health resources, codeveloped with students, tailored to sociocultural contexts, and integrated with mainstream in-person services at key touchpoints can significantly strengthen support systems and address the diverse needs of students in IHEs. To ensure sustained impact, this approach must be supported by coordinated national leadership, collaboratively evolved guidelines, shared financing mechanisms, including public-private partnerships, and long-term implementation–effectiveness research to address persistent gaps in student mental health care across India's higher education landscape.

Acknowledgments

This paper is based on the Indian Council of Medical Research (ICMR) funded project. ND, AG, and PV are ICMR staff members. The opinions expressed herein are those of the authors and do not necessarily represent the position of the ICMR.

Funding

The publication of this article was supported by funding from ICMR for an ongoing project at NIMHANS, Bengaluru.

Authors' Contributions

SM and TKS conceptualized the idea and wrote the initial draft of the manuscript. All the authors contributed to developing the ideas further and helped in arriving at the revised version of the manuscript.

Conflicts of Interest

None declared.

References

1. Karyotaki E, Cuijpers P, Albor Y, Alonso J, Auerbach RP, Bantjes J, et al. Sources of stress and their associations with mental disorders among college students: results of the World Health Organization World Mental Health Surveys International College Student Initiative. Front Psychol 2020;11:1759 [FREE Full text] [doi: 10.3389/fpsyg.2020.01759] [Medline: 32849042]
2. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
3. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization world mental health surveys. Psychol Med 2016;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
4. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]

XSL•FO
RenderX

5. Carrellas NW, Biederman J, Uchida M. How prevalent and morbid are subthreshold manifestations of major depression in adolescents? A literature review. J Affect Disord 2017;210:166-173. [doi: 10.1016/j.jad.2016.12.037] [Medline: 28049101]
6. Backhaus I, Varela AR, Khoo S, Siefken K, Crozier A, Begotaraj E, et al. Associations between social capital and depressive symptoms among college students in 12 countries: results of a cross-national study. Front Psychol 2020;11:644 [FREE Full text] [doi: 10.3389/fpsyg.2020.00644] [Medline: 32411038]
7. Suicide worldwide in 2019. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789240026643 [accessed 2025-10-02]
8. Accidental deaths and suicides in India 2022. National Crime Records Bureau. Ministry of Home Affairs, Government of India. 2022. URL: https://www.ncrb.gov.in/accidental-deaths-suicides-in-india-adsi.html [accessed 2025-05-25]
9. Yadav S, Cunningham SA, Bhandari P, Mishra US, Aditi A, Yadav R. Changing pattern of suicide deaths in India. Lancet Reg Health Southeast Asia 2023;16:100265 [FREE Full text] [doi: 10.1016/j.lansea.2023.100265] [Medline: 37649643]
10. Student suicides: an epidemic sweeping India. IC3 Institute. 2023. URL: https://ic3institute.org/wp-content/uploads/2023/09/Student-suicides.pdf [accessed 2025-05-26]
11. Amudhan S, Gururaj G, Varghese M, Benegal V, Rao GN, Sheehan DV, et al. A population-based analysis of suicidality and its correlates: findings from the National Mental Health Survey of India, 2015-16. Lancet Psychiatry 2020;7(1):41-51. [doi: 10.1016/S2215-0366(19)30404-3] [Medline: 31826821]
12. Cherian AV, Armstrong G, Sobhana H, Haregu T, Deuri SP, Bhat SU, et al. Mental health, suicidality, health, and social indicators among college students across nine states in India. Indian J Psychol Med 2025;47(3):253-260 [FREE Full text] [doi: 10.1177/02537176241244775] [Medline: 39564264]
13. Cuijpers P, Miguel C, Ciharova M, Aalten P, Batelaan N, Salemink E, et al. Prevention and treatment of mental health and psychosocial problems in college students: an umbrella review of meta-analyses. Clin Psychol Sci Pract 2021;28(3):229-244. [doi: 10.1037/cps0000030]
14. Jayasankar P, Manjunatha N, Rao GN, Gururaj G, Varghese M, Benegal V, NMHS India National Collaborator Group. Epidemiology of common mental disorders: results from "National Mental Health Survey" of India, 2016. Indian J Psychiatry 2022;64(1):13-19 [FREE Full text] [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_865_21] [Medline: 35400745]
15. The State of the World's Children 2021: on my mind — promoting, protecting, and caring for children's mental health. United Nations Children's Fund (UNICEF). 2021. URL: https://www.unicef.org/lac/media/28726/file/SOWC2021-full-report-English.pdf [accessed 2025-05-25]
16. Bruffaerts R, Mortier P, Auerbach RP, Alonso J, Hermosillo De la Torre AE, Cuijpers P, WHO WMH-ICS Collaborators. Lifetime and 12-month treatment for mental disorders and suicidal thoughts and behaviors among first year college students. Int J Methods Psychiatr Res 2019;28(2):e1764 [FREE Full text] [doi: 10.1002/mpr.1764] [Medline: 30663193]
17. Pedrelli P, Nyer M, Yeung A, Zulauf C, Wilens T. College students: mental health problems and treatment considerations. Acad Psychiatry 2015;39(5):503-511 [FREE Full text] [doi: 10.1007/s40596-014-0205-9] [Medline: 25142250]
18. Osborn TG, Li S, Saunders R, Fonagy P. University students' use of mental health services: a systematic review and meta-analysis. Int J Ment Health Syst 2022;16(1):57 [FREE Full text] [doi: 10.1186/s13033-022-00569-0] [Medline: 36527036]
19. Patel V, Xiao S, Chen H, Hanna F, Jotheeswaran AT, Luo D, et al. The magnitude of and health system responses to the mental health treatment gap in adults in India and China. Lancet 2016;388(10063):3074-3084. [doi: 10.1016/S0140-6736(16)00160-4] [Medline: 27209149]
20. Lattie EG, Lipson SK, Eisenberg D. Technology and college student mental health: challenges and opportunities. Front Psychiatry 2019;10:246 [FREE Full text] [doi: 10.3389/fpsyt.2019.00246] [Medline: 31037061]
21. D'Adamo L, Paraboschi L, Grammer AC, Fennig M, Graham AK, Yaeger LH, et al. Reach and uptake of digital mental health interventions based on cognitive-behavioral therapy for college students: a systematic review. J Behav Cogn Ther 2023;33(2):97-117 [FREE Full text] [doi: 10.1016/j.jbct.2023.05.002] [Medline: 37724304]
22. Gask L, Bower P, Lamb J, Burroughs H, Chew-Graham C, Edwards S, et al. Improving access to psychosocial interventions for common mental health problems in the United Kingdom: narrative review and development of a conceptual model for complex interventions. BMC Health Serv Res 2012;12:249 [FREE Full text] [doi: 10.1186/1472-6963-12-249] [Medline: 22889290]
23. Sanghvi P, Mehrotra S, Das SK, Sharma MK. Barriers to seek professional help for mental health scale: development and psychometric validation. Indian J Psychol Med 2024:02537176241256480. [doi: 10.1177/02537176241256480] [Medline: 39564249]
24. Choudhary S, Dutt S, Torous J, Mehta UM. Bridging the mental health gap: the scope, challenges, and future of digital psychiatry in India. Asian J Psychiatr 2024;101:104215. [doi: 10.1016/j.ajp.2024.104215] [Medline: 39243661]
25. Internet in India 2024. Internet and Mobile Association of India. URL: https://www.iamai.in/sites/default/files/research/Kantar_%20IAMAI%20report_2024_.pdf [accessed 2025-05-25]
26. Singh SK, Singh V. Internet diffusion in India: a study based on growth curve modelling. Manag Res Pract 2023;15(2):29-42 [FREE Full text]
27. Aggarwal NK. Applying mobile technologies to mental health service delivery in South Asia. Asian J Psychiatr 2012;5(3):225-230. [doi: 10.1016/j.ajp.2011.12.009] [Medline: 22981050]

XSL•FO
RenderX

28. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010;192(S11):S22-S26. [doi: 10.5694/j.1326-5377.2010.tb03688.x] [Medline: 20528703]
29. Burns J, Birrell E. Enhancing early engagement with mental health services by young people. Psychol Res Behav Manag 2014;7:303-312 [FREE Full text] [doi: 10.2147/PRBM.S49151] [Medline: 25473320]
30. March S, Day J, Ritchie G, Rowe A, Gough J, Hall T, et al. Attitudes toward e-mental health services in a community sample of adults: online survey. J Med Internet Res 2018;20(2):e59 [FREE Full text] [doi: 10.2196/jmir.9109] [Medline: 29459357]
31. Topooco N, Fowler LA, Fitzsimmons-Craft EE, DePietro B, Vázquez MM, Firebaugh M, et al. Digital interventions to address mental health needs in colleges: perspectives of student stakeholders. Internet Interv 2022;28:100528 [FREE Full text] [doi: 10.1016/j.invent.2022.100528] [Medline: 35378846]
32. Abrams Z. How well is telepsychology working? Mon Psychol 2020;51(5):46 [FREE Full text] [doi: 10.1007/978-3-030-39903-0_301961]
33. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
34. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
35. Harith S, Backhaus I, Mohbin N, Ngo HT, Khoo S. Effectiveness of digital mental health interventions for university students: an umbrella review. PeerJ 2022;10:e13111 [FREE Full text] [doi: 10.7717/peerj.13111] [Medline: 35382010]
36. Madrid-Cagigal A, Kealy C, Potts C, Mulvenna MD, Byrne M, Barry MM, et al. Digital mental health interventions for university students with mental health difficulties: a systematic review and meta-analysis. Early Interv Psychiatry 2025;19(3):e70017. [doi: 10.1111/eip.70017] [Medline: 40033658]
37. Conley CS, Durlak JA, Kirsch AC. A Meta-analysis of universal mental health prevention programs for higher education students. Prev Sci 2015;16(4):487-507. [doi: 10.1007/s11121-015-0543-1] [Medline: 25744536]
38. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]
39. Vaingankar JA, van Dam RM, Samari E, Chang S, Seow E, Chua YC, et al. Social media-driven routes to positive mental health among youth: qualitative enquiry and concept mapping study. JMIR Pediatr Parent 2022;5(1):e32758 [FREE Full text] [doi: 10.2196/32758] [Medline: 35254285]
40. Tam MT, Wu JM, Zhang CC, Pawliuk C, Robillard JM. A systematic review of the impacts of media mental health awareness campaigns on young people. Health Promot Pract 2024;25(5):907-920 [FREE Full text] [doi: 10.1177/15248399241232646] [Medline: 38468568]
41. Haas A, Koestner B, Rosenberg J, Moore D, Garlow SJ, Sedway J, et al. An interactive web-based method of outreach to college students at risk for suicide. J Am Coll Health 2008;57(1):15-22. [doi: 10.3200/JACH.57.1.15-22] [Medline: 18682341]
42. Kamel Boulos M, Giustini D, Wheeler S. Instagram and WhatsApp in health and healthcare: an overview. Future Internet 2016;8(3):37 [FREE Full text] [doi: 10.3390/fi8030037]
43. Xu Z, Huang F, Kösters M, Staiger T, Becker T, Thornicroft G, et al. Effectiveness of interventions to promote help-seeking for mental health problems: systematic review and meta-analysis. Psychol Med 2018;48(16):2658-2667. [doi: 10.1017/S0033291718001265] [Medline: 29852885]
44. Johnson JA, Sanghvi P, Mehrotra S. Technology-based interventions to improve help-seeking for mental health concerns: a systematic review. Indian J Psychol Med 2022;44(4):332-340 [FREE Full text] [doi: 10.1177/02537176211034578] [Medline: 35949632]
45. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
46. Wadhwa M. Digital Mental Health Initiatives in India. ICT India Working Paper, No. 60. Econstor. 2021. URL: https://www.econstor.eu/bitstream/10419/249849/1/ICT-India-Working-Paper-60.pdf [accessed 2025-05-25]
47. Maroju RG, Choudhari SG, Shaikh MK, Borkar SK, Mendhe H. Role of telemedicine and digital technology in public health in india: a narrative review. Cureus 2023;15(3):e35986 [FREE Full text] [doi: 10.7759/cureus.35986] [Medline: 37050980]
48. Chandra PS, Sowmya HR, Mehrotra S, Duggal M. 'SMS' for mental health—feasibility and acceptability of using text messages for mental health promotion among young women from urban low income settings in India. Asian J Psychiatr 2014;11:59-64. [doi: 10.1016/j.ajp.2014.06.008] [Medline: 25453699]
49. Maulik PK, Kallakuri S, Devarapalli S, Vadlamani VK, Jha V, Patel A. Increasing use of mental health services in remote areas using mobile technology: a pre-post evaluation of the SMART Mental Health project in rural India. J Glob Health 2017;7(1):010408 [FREE Full text] [doi: 10.7189/jogh.07.010408] [Medline: 28400954]

XSL•FO
RenderX

50. Tewari A, Kallakuri S, Devarapalli S, Jha V, Patel A, Maulik PK. Process evaluation of the Systematic Medical Appraisal, Referral and Treatment (SMART) mental health project in rural India. BMC Psychiatry 2017;17(1):385 [FREE Full text] [doi: 10.1186/s12888-017-1525-6] [Medline: 29202773]
51. Choubisa R, Singh K. Development and validation of a web-delivered positive psychological intervention in an Indian Milieu: lessons from a limited pilot randomized controlled trial. Cogent Psychol 2018 Jul 04;5(1):1-28. [doi: 10.1080/23311908.2018.1488512]
52. Kanuri N, Arora P, Talluru S, Colaco B, Dutta R, Rawat A, et al. Examining the initial usability, acceptability and feasibility of a digital mental health intervention for college students in India. Int J Psychol 2020;55(4):657-673 [FREE Full text] [doi: 10.1002/ijop.12640] [Medline: 31867730]
53. Mehrotra S, Sudhir P, Rao G, Thirthalli J, Srikanth TK. Development and pilot testing of an internet-based self-help intervention for depression for indian users. Behav Sci (Basel) 2018;8(4):36 [FREE Full text] [doi: 10.3390/bs8040036] [Medline: 29565278]
54. Karishiddimath A, Francis A, Harikumar H, Sengupta P, Srikanth TK, Mehrotra S. Preliminary evaluation of an indigenous mental health app for indian users. Cureus 2025;17(3):e80346. [doi: 10.7759/cureus.80346] [Medline: 40206925]
55. Sudarshan S, Mehrotra S, Thirthalli J. Effectiveness of blended internet-based self-help and face-to-face intervention for depression: a pilot study from india. Indian J Psychol Med 2025;47(2):142-153 [FREE Full text] [doi: 10.1177/02537176241238289] [Medline: 39564242]
56. Sanghvi P, Mehrotra S, Sharma M. "ReachOut": pilot evaluation of a help-seeking intervention for common mental health concerns among distressed non-treatment-seeking young adults. Cureus 2024;16(2):e54324 [FREE Full text] [doi: 10.7759/cureus.54324] [Medline: 38371429]
57. Kumar CN, Chand PK, Manjunatha N, Math SB, Shashidhara HN, Basavaraju V, et al. Impact evaluation of VKN-NIMHANS-ECHO model of capacity building for mental health and addiction: methodology of two randomized controlled trials. Indian J Psychol Med 2020;42(6 Suppl):S80-S86 [FREE Full text] [doi: 10.1177/0253717620969066] [Medline: 33487808]
58. Sagar R, Singh S. National Tele-Mental Health Program in India: a step towards mental health care for all? Indian J Psychiatry 2022;64(2):117-119 [FREE Full text] [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_145_22] [Medline: 35494321]
59. Bear HA, Ayala Nunes L, Ramos G, Manchanda T, Fernandes B, Chabursky S, et al. The acceptability, engagement, and feasibility of mental health apps for marginalized and underserved young people: systematic review and qualitative study. J Med Internet Res 2024;26:e48964 [FREE Full text] [doi: 10.2196/48964] [Medline: 39078699]
60. Mehrotra S, Tripathi R. Recent developments in the use of smartphone interventions for mental health. Curr Opin Psychiatry 2018;31(5):379-388. [doi: 10.1097/YCO.0000000000000439] [Medline: 29985179]
61. Yorgason JB, Linville D, Zitzman B. Mental health among college students: do those who need services know about and use them? J Am Coll Health 2008;57(2):173-181. [doi: 10.3200/JACH.57.2.173-182] [Medline: 18809534]
62. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
63. Fullmer L, Fleming A, Green K. Perceptions of help-seeking barriers among college students with mental health disabilities. Rehabil Couns Educ J 2021;10(1):1-10. [doi: 10.52017/001c.24870]
64. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
65. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interventions 2014;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
66. Özer Ö, Köksal B, Altinok A. Understanding university students' attitudes and preferences for internet-based mental health interventions. Internet Interv 2024;35:100722 [FREE Full text] [doi: 10.1016/j.invent.2024.100722] [Medline: 38356613]
67. Yuen EK, Gangi CE, Barakat K, Harrison F. College students' utilization of the internet to search for mental health information: effects on mental health literacy, stigma, and help-seeking. J Am Coll Health 2025;73(3):961-971. [doi: 10.1080/07448481.2024.2404948] [Medline: 39303085]
68. Hadler NL, Bu P, Winkler A, Alexander AW. College student perspectives of telemental health: a review of the recent literature. Curr Psychiatry Rep 2021;23(2):6 [FREE Full text] [doi: 10.1007/s11920-020-01215-7] [Medline: 33404975]
69. Conley CS, Durlak JA, Shapiro JB, Kirsch AC, Zahniser E. A meta-analysis of the impact of universal and indicated preventive technology-delivered interventions for higher education students. Prev Sci 2016;17(6):659-678. [doi: 10.1007/s11121-016-0662-3] [Medline: 27225631]
70. Farrer L, Gulliver A, Chan JKY, Batterham PJ, Reynolds J, Calear A, et al. Technology-based interventions for mental health in tertiary students: systematic review. J Med Internet Res 2013;15(5):e101 [FREE Full text] [doi: 10.2196/jmir.2639] [Medline: 23711740]
71. Braciszewski JM. Digital technology for suicide prevention. Adv Psychiatry Behav Health 2021;1(1):53-65. [doi: 10.1016/j.ypsc.2021.05.008]

XSL•FO
RenderX

72. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
73. Abelson S, Eisenberg D, Johnston A, Lipson S, Liu M, Ogden S, et al. Digital mental health interventions at colleges and universities: understanding the need, assessing the evidence, and identifying steps forward. The Hope Center, Temple University. 2024. URL: https://hope.temple.edu/digital-mental-health-interventions-highered [accessed 2025-05-25]
74. Toscos T, Carpenter M, Drouin M, Roebuck A, Kerrigan C, Mirro M. College students' experiences with, and willingness to use, different types of telemental health resources: do gender, depression/anxiety, or stress levels matter. Telemed J E Health 2018;24(12):998-1005. [doi: 10.1089/tmj.2017.0243] [Medline: 29658826]
75. Kern A, Hong V, Song J, Lipson SK, Eisenberg D. Mental health apps in a college setting: openness, usage, and attitudes. Mhealth 2018;4:20 [FREE Full text] [doi: 10.21037/mhealth.2018.06.01] [Medline: 30050916]
76. Santucci LC, McHugh RK, Elkins RM, Schechter B, Ross MS, Landa CE, et al. Pilot implementation of computerized cognitive behavioral therapy in a university health setting. Adm Policy Ment Health 2014;41(4):514-521. [doi: 10.1007/s10488-013-0488-2] [Medline: 23592231]
77. Levin ME, Pistorello J, Hayes SC, Seeley JR, Levin C. Feasibility of an acceptance and commitment therapy adjunctive web-based program for counseling centers. J Couns Psychol 2015;62(3):529-536. [doi: 10.1037/cou0000083] [Medline: 25961755]
78. Montagni I, Tzourio C, Cousin T, Sagara JA, Bada-Alonzi J, Horgan A. Mental health-related digital use by university students: a systematic review. Telemed J E Health 2020;26(2):131-146. [doi: 10.1089/tmj.2018.0316] [Medline: 30888256]
79. AISHE final report. AISHE, Ministry of Education, Government of India. 2022. URL: https://aishe.gov.in/aishe-final-report/ [accessed 2025-05-25]
80. Upadhyaya R. Demand for grants 2024-25 analysis: education. PRS Legislative Research, Institute for Policy Research Studies. 2024. URL: https://prsindia.org/files/budget/budget_parliament/2024/DFG_2024-25_Analysis_Education.pdf [accessed 2025-05-25]
81. National Education Policy 2020. Ministry of Human Resource Development (now Ministry of Education), Government of India. 2020. URL: https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English.pdf [accessed 2025-05-25]
82. Aithal PS, Prabhu S, Aithal S. Future of higher education through technology prediction and forecasting. Poornaprajna Int J Manag Educ Soc Sci 2024;1(1):1-50. [doi: 10.2139/ssrn.4901474]
83. Kaur R. Digitalization of higher education: issues and challenges. Int J Appl Res 2019;5(SP4):117-120 [FREE Full text]
84. India employment report 2024: youth employment, education and skills. International Labour Organization. 2024. URL: https://www.ilo.org/publications/india-employment-report-2024-youth-employment-education-and-skills [accessed 2025-05-25]
85. George AS. The Race to Success: A Study of the NEET and JEE Exams' Impact on Students' Lives. Tamilnadu: PU Publications; 2023.
86. Wasil AR, Malhotra T, Nandakumar N, Tuteja N, DeRubeis RJ, Stewart RE, et al. Improving mental health on college campuses: perspectives of Indian college students. Behav Ther 2022;53(2):348-364. [doi: 10.1016/j.beth.2021.09.004] [Medline: 35227409]
87. Guidelines for promotion of physical fitness, sports, students' health, welfare, psychological and emotional well-being at higher educational institutions of India. University Grants Commission. 2023. URL: https://www.ugc.gov.in/pdfnews/7223595_Guidelines-for-Promotion-of-Physical-Fitness-Sports-Students-Health-Welfare-Psychological-and-Emotional-Well-Being-at-Higher-Educational-Institutions-of-India.pdf [accessed 2025-05-25]
88. Aithal PS, Aithal S. Implementation strategies of higher education part of National Education Policy 2020 of India towards achieving its objectives. Int J Manag Technol Soc Sci 2020;5(2):283-325. [doi: 10.2139/ssrn.3741425]
89. National suicide prevention strategy. Ministry of Health and Family Welfare, Government of India. 2022. URL: https://cdnbbsr.s3waas.gov.in/s3bc88bb2c377261b0336312714b96fcfd/uploads/2025/04/20250507983470082.pdf [accessed 2025-05-25]
90. The Digital Personal Data Protection Act. Government of India. 2025. URL: https://indiankanoon.org/doc/3510545/ [accessed 2025-05-25]
91. Ethical guidelines for application of artificial intelligence in biomedical research and healthcare. Indian Council of Medical Research. 2023. URL: https://www.icmr.gov.in/icmrobject/custom_data/pdf/Ethical-guidelines/Ethical_Guidelines_AI_Healthcare_2023.pdf [accessed 2025-05-25]
92. Cho H, Jung K, Eisenberg D, King C, Zheng K. Barriers to digital mental health services among college students. Stud Health Technol Inform 2025;329:1261-1265. [doi: 10.3233/SHTI251041] [Medline: 40776059]
93. Taylor ME, Liu M, Abelson S, Eisenberg D, Lipson SK, Schueller SM. The reach, effectiveness, adoption, implementation, and maintenance of digital mental health interventions for college students: a systematic review. Curr Psychiatry Rep 2024;26(12):683-693 [FREE Full text] [doi: 10.1007/s11920-024-01545-w] [Medline: 39392547]
94. Wani C, McCann L, Lennon M, Radu C. Digital mental health interventions for adolescents in low- and middle-income countries: scoping review. J Med Internet Res 2024;26:e51376 [FREE Full text] [doi: 10.2196/51376] [Medline: 39471371]

XSL•FO
RenderX

95. Kodish T, Schueller SM, Lau AS. Barriers and strategies to improve digital mental health intervention uptake among college students of color: a modified Delphi study. J Behav Cogn Ther 2023;33(1):10-23. [doi: 10.1016/j.jbct.2022.12.002]
96. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
97. Oti O, Pitt I. Online mental health interventions designed for students in higher education: a user-centered perspective. Internet Interv 2021;26:100468 [FREE Full text] [doi: 10.1016/j.invent.2021.100468] [Medline: 34703772]
98. Mehrotra S, Tripathi R, Sengupta P, Karishiddimath A, Francis A, Sharma P, et al. Evaluating characteristics and quality of mental health apps available in app stores for indian users: systematic app search and review. JMIR Mhealth Uhealth 2025;13:e79238 [FREE Full text] [doi: 10.2196/79238] [Medline: 41004798]
99. Pillutla R. Delving into Digital Mental Health: Part 1. IMHO. 2024. URL: https://cmhlp.org/wp-content/uploads/2024/04/Digital-Mental-Health-Part-1.pdf [accessed 2025-05-26]
100. Singh Sethi MI, Kumar RC, Manjunatha N, Naveen Kumar C, Math SB. Mental health apps in India: regulatory landscape and future directions. BJPsych Int 2025;22(1):2-5. [doi: 10.1192/bji.2024.20] [Medline: 40290478]
101. Kumar S, Mehrotra S. Free mobile apps on depression for Indian users: a brief overview and critique. Asian J Psychiatr 2017;28:124-130. [doi: 10.1016/j.ajp.2017.03.031] [Medline: 28784364]
102. Smith KA, Blease C, Faurholt-Jepsen M, Firth J, Van Daele T, Moreno C, et al. Digital mental health: challenges and next steps. BMJ Ment Health 2023;26(1):e300670 [FREE Full text] [doi: 10.1136/bmjment-2023-300670] [Medline: 37197797]
103. Zhu S, Wang Y, Hu Y. Facilitators and barriers to digital mental health interventions for depression, anxiety, and stress in adolescents and young adults: scoping review. J Med Internet Res 2025;27:e62870 [FREE Full text] [doi: 10.2196/62870] [Medline: 40127430]
104. Mehrotra S, Kumar S, Sudhir P, Rao GN, Thirthalli J, Gandotra A. Unguided mental health self-help apps: reflections on challenges through a clinician's lens. Indian J Psychol Med 2017;39(5):707-711 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_151_17] [Medline: 29200577]
105. Banker DV, Bhal KT. Creating world class universities: roles and responsibilities for academic leaders in India. Educ Manag Admin Leadersh 2019;48(3):570-590. [doi: 10.1177/1741143218822776]
106. Mehrotra S. Indian higher education and youth mental health: challenges and opportunities. J Glob Health 2020;10(2):020307 [FREE Full text] [doi: 10.7189/jogh.10.020307] [Medline: 33110511]
107. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
108. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024;24(1):243 [FREE Full text] [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
109. Ben-Zeev D, Razzano LA, Pashka NJ, Levin CE. Cost of mHealth versus clinic-based care for serious mental illness: same effects, half the price tag. Psychiatr Serv 2021;72(4):448-451. [doi: 10.1176/appi.ps.202000349] [Medline: 33557599]
110. Joshi U, Naslund JA, Anand A, Tugnawat D, Vishwakarma R, Bhan A, et al. Assessing costs of developing a digital program for training community health workers to deliver treatment for depression: a case study in rural India. Psychiatry Res 2022;307:114299 [FREE Full text] [doi: 10.1016/j.psychres.2021.114299] [Medline: 34871874]
111. Vohra S, Grover S, Prakash A, Khurana D. Public-private partnership model in digital mental healthcare domain: way forward for india and the developing world. J Technol Behav Sci 2024;9(3):526-531. [doi: 10.1007/s41347-024-00412-y]
112. National Dialysis Program under National Health Mission. Ministry of Health and Family Welfare, Government of India. 2016. URL: https://www.nhm.gov.in/images/pdf/NHM/NHM-Guidelines/Operational_Guidelines_for_Dialysis_Programme.pdf [accessed 2025-10-11]
113. De Costa A, Vora KS, Ryan K, Sankara Raman P, Santacatterina M, Mavalankar D. The state-led large scale public private partnership 'Chiranjeevi Program' to increase access to institutional delivery among poor women in Gujarat, India: how has it done? What can we learn? PLoS One 2014;9(5):e95704 [FREE Full text] [doi: 10.1371/journal.pone.0095704] [Medline: 24787692]
114. Public private partnerships. Apollo TeleHealth. 2025. URL: https://www.apollotelehealth.com/public-private-partnerships/ [accessed 2025-10-11]
115. Sukdeb Saha v. State of Andhra Pradesh. 2025 INSC 893. Supreme Court Observer (SCO). URL: https://www.scobserver.in/wp-content/uploads/2025/07/SCOLR_judgement_Sukdeb-Saha-v-State-of-AP.pdf [accessed 2025-10-11]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
DMHI: Digital Mental Health Intervention
DPDP: Digital Personal Data Protection
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)

XSL•FO
RenderX

IHEs: Institutes of Higher Education
LMIC: low- and middle-income country
NEP: National Education Policy
PPP: Public-Private-Partnership
WHO: World Health Organization

Edited by E Mensah; submitted 01.06.25; peer-reviewed by A Tai, SK Geale; comments to author 26.09.25; accepted 06.11.25; published 21.11.25.

Please cite as:
Mehrotra S, Srikanth TK, Dahiya N, Verma P, Rao GN, Sanghvi P, Grover A, Sagar R
Technology for Mental Health: Reflections on Scope and Future Directions in Institutes of Higher Education in India
Online J Public Health Inform 2025;17:e78065
URL: https://ojphi.jmir.org/2025/1/e78065
doi:10.2196/78065
PMID:41271205

©Seema Mehrotra, TK Srikanth, Neha Dahiya, Pulkit Verma, Girish N Rao, Prachi Sanghvi, Ashoo Grover, Rajesh Sagar. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 21.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Imposters, Bots, and Other Threats to Data Integrity in Online Research: Scoping Review of the Literature and Recommendations for Best Practices

Isabella B Strickland[1]; Amy K Ferketich[1,2], PhD; Alayna P Tackett[2,3], PhD; Joanne G Patterson[1], PhD; Nicholas J K Breitborde[4], PhD; Jade Davis[1]; Megan Roberts[1,2], PhD

[1]College of Public Health, The Ohio State University, 1841 Neil Ave, Columbus, OH, United States

[2]Center for Tobacco Research, The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH, United States

[3]Division of Medical Oncology, Department of Internal Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States

[4]Psychiatry and Behavioral Health, College of Medicine, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Megan Roberts, PhD
College of Public Health, The Ohio State University, 1841 Neil Ave, Columbus, OH, United States

Abstract

Background: Threats to data integrity have always existed in online human subjects research, but it appears these threats have become more common and more advanced in recent years. Researchers have proposed various techniques to address satisficers, repeat participants, bots, and fraudulent participants; yet, no synthesis of this literature has been conducted.

Objective: This study undertakes a scoping review of recent methods and ethical considerations for addressing threats to data integrity in online research.

Methods: A PubMed search was used to identify 90 articles published from 2020 to 2024 that were written in English, that discussed online human subjects research, and that had at least one paragraph dedicated to discussing threats to online data integrity.

Results: We cataloged 16 types of techniques for addressing threats to online data integrity. Techniques to authenticate personal information (eg, videoconferencing and mailing incentives to a physical address) appear to be very effective at deterring or identifying fraudulent participants. Yet such techniques also come with ethical considerations, including participant burden and increased threats to privacy. Other techniques, such as Completely Automated Public Turing test to tell Computers and Humans Apart (reCAPTCHA; Google LLC), scores, and checking IP addresses, although very common, were also deemed by several researchers as no longer sufficient protections against advanced threats to data integrity.

Conclusions: Overall, this review demonstrates the importance of shifting online research protocols as bots and fraudulent participants become more sophisticated.

(Online J Public Health Inform 2025;17:e70926) doi:10.2196/70926

KEYWORDS

review; fraud; data integrity; bots; online data collection; PRISMA

Introduction

Recent years have witnessed a shift in research protocols, with many studies that were previously conducted in-person being moved online [1]. This shift has had several benefits for researchers in terms of easier sampling, broader reach, and better access to historically marginalized populations [2,3]. However, the shift has also ushered in critical concerns about threats to data integrity. While concerns about data integrity have always existed, even with in-person studies, there has been a notable increase in the number and types of threats to data integrity in online studies since the COVID-19 pandemic [4,5].

As outlined in Table 1, the types of threats to data integrity in online research come in several forms. First, satisficers (also known as speeders, straightliners, and cheaters) are individuals who rush through surveys with little care for the accuracy or thoroughness of their responses. While satisficers also exist with study protocols administered in-person or by mail [6,7], this threat is challenging to monitor online. Next, there are repeat participants (also known as duplicate participants). These are individuals who complete screener surveys or study protocols multiple times. Motivations for repeat participation vary: some people may be curious about what happens if they complete the survey with different answers [8]. Others may try to complete the survey multiple times to get extra compensation [9].

XSL•FO
RenderX

Whatever the motivation, this behavior can have serious consequences for data integrity and research findings [10]. And, again, it is often easier for individuals to engage in these behaviors with online research studies.

Table . Types of threats to data integrity in online research.

Type of threat	Other terms	Definition
Satisficers	Cheaters, straightliners, speeders, and careless participants	Inattentive participants who speed through surveys often not paying attention to questions and responding thoughtlessly
Repeat participants	Duplicate participants	Participants who attempt to complete a study more than once out of curiosity or a desire for additional remuneration
Bots	Chatbots and artificial intelligence respondents	Computer algorithms deployed on studies in order to gain compensation without human effort for completion
Fraudulent participants	Imposters, scammers, bad actors, and lying participants	Participants who lie about their identity or otherwise attempt to deceive researchers often with the intent of gaining study compensation

Another growing threat to data integrity is bots (also known as chatbots and artificial intelligence respondents). Bots are automated computer programs that people can create to randomly and methodically complete online surveys, usually numerous times, to gain compensation without having to take the time and effort to manually complete the survey [11]. Tools such as a Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA; or the newer version, reCAPTCHA; Google LLC) can help prevent bots from infiltrating surveys; however, more advanced bots can bypass these measures [12]. Bots can very quickly complete surveys and, if they are not properly blocked, can compromise results and force time-consuming and expensive relaunches of research projects [13,14].

Finally, there are fraudulent participants (also known as imposter participants, scammers, and lying participants) [4,5,15]. Unlike bots, fraudulent participants are real people who complete a study protocol. However, they lie about themselves to qualify for study participation. For example, Pellicano et al [4] describe a situation where fraudulent participants posed as either people with autism or parents of children with autism during online, qualitative interviews. In this particular example, several clues aroused the researcher's suspicion, such as keeping cameras off, inconsistent responses between prescreening and the interview, similarities in voices and mannerisms across interviews, and repeated inquiries about payments. Fraudulent participants have been detected in many domains of research but seem, concerningly, to have the largest impact on research on small populations that are often historically minoritized or otherwise vulnerable [10]. One study interviewing research participants found that, on average, 55% of participants who had participated in some sort of research fraud reported fabricating information to qualify for studies [16]. Several studies have found that these participants often respond differently than authentic participants, potentially influencing the results of research studies or weakening the effects detected [17-19].

While many researchers have published concerns or potential solutions to these various threats to data integrity, there has not yet been a review or synthesis of the literature. This research gap makes it difficult and time-consuming for researchers designing new online studies to decide on best practices. Due to the research interests of the authors, we were particularly motivated to identify research methods being used to address threats to online data integrity in the medical and public health domains. Therefore, we used PubMed to conduct a scoping review of methods that address contemporary threats to online data integrity, with keywords that focus on bots and fraudulent participants. Our objectives were to catalog and evaluate the most common research methods used to address these threats and discuss the ethical considerations raised about the techniques. Ultimately, this review aims to expand and centralize knowledge on addressing threats to data integrity in online studies, with the goal of aiding researchers in developing robust online methodologies.

Methods

Search Strategy

This search was conducted using Covidence (Veritas Health Innovation), an online tool used to conduct and organize literature reviews. We searched PubMed for the terms "fraud* OR imposter* OR scam* OR bot OR bots." To be eligible for review, papers needed to be published in or after 2020 because we were interested in methods for addressing the recent threats to data integrity that have emerged since COVID-19. Additional eligibility criteria were: discussing an online study, using human participants, being written in English, and having at least one paragraph dedicated to discussing threats to data integrity.

Our PubMed search yielded an initial 10,681 publications. After eliminating 189 duplicate texts using Covidence, 10,492 publications remained to be screened. Our first phase of screening checked all abstracts and eliminated those not pertaining to online research. Of the 10,492 publications, 196 were retained. In the second phase of screening, the full article was checked for inclusion criteria. Ultimately, 90 articles met criteria for inclusion in this review [2-4,11,13,15,20-103]. This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines.

XSL•FO
RenderX

Data Extraction

A codebook was developed through an interrogative process. First, a codebook was created a priori according to our research questions. Additional codes and subcodes were added through an inductive process. Two team members (IBS and JD) independently reviewed and coded each of the 90 articles [2-4,11,13,15,20-103]. A senior team member (MR) reviewed their interrater agreement and resolved all discrepancies.

The following information was extracted and coded from the final 90 articles [2-4,11,13,15,20-103]: article type (eg, original research and commentary), type of data collection (eg, qualitative and quantitative), methodology of the study (eg, survey and qualitative interview), country where the study was conducted, recruitment methods (eg, social media and survey service platform), type of suspected threat to data integrity (eg, bots and fraudulent participants), the estimated prevalence of compromised data, and techniques mentioned for addressing threats to data integrity (eg, authenticating personal information and attention checks).

Techniques to address threats to data integrity were additionally sorted into 3 categories. “Very effective” techniques were those that authors of the reviewed articles, especially in the most recent publications, deemed to be successful at identifying poor-quality data, such as bots and fraudulent participants. “Somewhat effective” techniques were those considered capable of identifying a proportion of poor-quality data but that had drawbacks preventing them from being used alone. “No longer effective” techniques were those deemed by the authors of the reviewed articles as being no longer sufficient in addressing threats to data integrity.

When synthesizing the data, we computed the most common types of threats to data integrity, the most common recruitment methods, and the estimated prevalence of threats to data integrity. We also narratively reviewed how authors discussed the adverse effects of threats to data integrity. Next, to address our study objectives, we described all the proposed techniques to address threats to data integrity that we uncovered in this review. Finally, we narratively reviewed how authors discussed the ethical considerations raised by the techniques.

Results

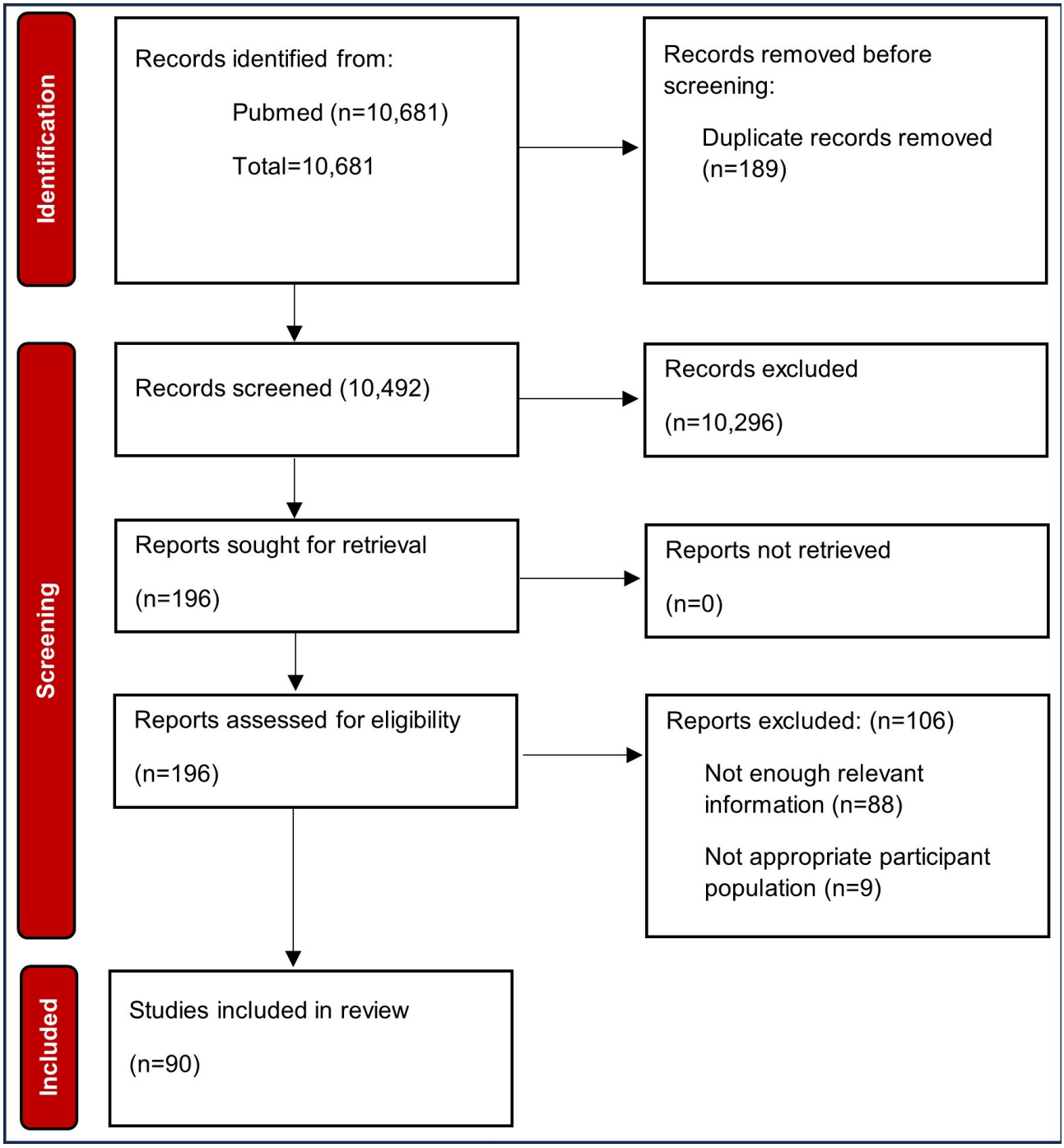
A total of 90 studies were included in the review (Figure 1). The most common type of threat to data integrity documented by researchers was bots (n=59) followed by fraudulent participants (n=51), repeat participants (n=42), and satisficing participants (n=17). The most common recruitment method was social media advertising, followed by using online survey service platforms, such as MTurk (Amazon Web Services, Inc; Table S1 in Multimedia Appendix 1 [2-4,11,13,15,20-103]). Many studies used more than one recruitment method.

The estimated prevalence of threats to data integrity ranged from approximately 1% to 99%. Implications for data validity and reliability were commonly discussed. Other adverse effects included the heavy, and often wasteful, use of resources needed to address fraud. For example, some researchers with a high prevalence of participant fraud described having to end their study and start over, wasting valuable time and resources. Some articles even discussed how dealing with high proportions of imposter participants can be difficult to handle emotionally as researchers. As one research team expressed after finding around 90% of their study participants to be fraudulent: “It is disheartening to encounter issues related to fraud during research. Our team experienced significant demoralization related to this occurrence” [20].

All of the articles provided information on methods to improve the integrity of data, either by (1) preventing the collection of poor-quality data in the first place or (2) identifying and removing poor-quality data if collected (or both; Table S2 in Multimedia Appendix 2 [2-4,11,13,15,20-103]). As cataloged in Table 2, we identified 16 techniques, representing a wide variety of methods. Techniques deemed to be “very effective” included authenticating personal information, such as requesting to see participants’ IDs over a video call, which eliminates the potential for bots and helps identify fraudulent participants. As another version of this technique, Hardesty et al [21] mailed the study incentives to participants’ street addresses (rather than sending the incentive electronically) because they observed that fraudulent participants were providing false addresses to meet geographically based eligibility criteria. A related technique deemed “very effective” was including background-related questions that could be easily answered by participants in the target population (or by partners in dyad research) but that are not widely known by other groups. For example, in a community study on narcolepsy, data were excluded from participants who reported unlikely symptomology [22]. A final “very effective” technique concerned data checking: cross-checking for inconsistent answers (eg, between screening and a baseline survey). Of note, we observed that these very effective techniques were used across a variety of study designs, including both quantitative and qualitative studies.

XSL•FO
RenderX

Figure 1. Flow chart of the review process for article selection.



XSL•FO
RenderX

Table . Techniques to address threats to data integrity in online research and their frequency of being mentioned across the 90 articles examined in this scoping review.

Effectiveness ranking and technique		Description	Example	Freq.[a]
Very effective				
	Authenticate personal information	Checking IDs, emails, addresses, zip codes, and phone numbers for authenticity. This could include using third-party services to verify identities, requiring video calls at enrollment (verification step), or mailing incentives to the provided street address.	"Include a preinterview briefing over videoconferencing or telephone to go through eligibility criteria and the consent process. Researchers could forewarn potential participants about this aspect in the consent form." [4]	48.9%
	Background-related questions	Including questions about information that would be easily answered by participants in target population but is not widely known by other groups.	"To reduce fraudulent responses, the study investigators added 4 military validation questions to confirm history of military service prior to the study survey. These questions were developed and piloted with service members and veterans of varying components and across branches." [23]	20.0%
	Cross-check inconsistent answers	Checking for inconsistent or contradictory answers across survey items to detect fraud or inattention.	"By identifying inconsistencies in data collected at screening and survey data, the team could identify potentially fraudulent or ineligible participants." [24]	51.1%
Somewhat effective				

XSL•FO
RenderX

Effectiveness ranking and technique		Description	Example	Freq.[a]
	Attention checks	Including survey questions that request specific answers or that may only have one reasonable answer. This screens for satisficing and basic logical reasoning.	"The attention checks consisted of the following: (1) embedded on the Grit scale—"Select 'Somewhat like me' for this statement," (2) embedded on the Beck Depression Inventory—"1 – Select this option," and (3) embedded on the Borderline Personality Inventory—"Select 'Yes' for this statement.'" [25]	33.3%
	Camera-on requirement	Requiring participants to turn on their camera, even if just for a moment, as often fraudulent participants will leave theirs off.	"Participants were not using their cameras for the Zoom sessions because they refused to turn on their camera or they stated there were internet issues." [26]	10.0%
	Check for dataset duplicates	Checking a dataset for duplicate names, emails, etc, across participants for duplicate replies.	"SAS programs were run to check the newly submitted record against all previous baseline questionnaires to check for duplicates of email addresses, mobile numbers, IP addresses, mailing addresses, social media handles, and preferred names." [24]	46.7%
	Post hoc testing	Conducting statistical analysis of data for unreasonable response patterns and notable outliers that may be indicative of fraud.	"Interactive visualization can improve data quality by facilitating the identification of issues such as missing data, outliers, duplicates, pattern or constraint violations, and data inconsistencies." [27]	16.7%
	Watchful of a large number of responses	Checking the timestamps on survey submissions. A flurry of responses or sign-ups can often be an indication of bots or fraud.	"Before launching the DIP, various indications of fraudulent activity were noted. These include…a rush of survey time stamps…found in the same 1 - to 15-min period." [28]	30.0%
	Screen for low response rates	Excluding data from participants who complete less than a certain percentage of a study, as they may be satisficing.	"Frequently examine the data for any patterns such as large blocks of blank question." [29]	8.9%
	Changing payment protocols	Intentionally not emphasizing and not automating participant payments. For example, not including the payment amount in recruitment materials.	"To maximize reach and limit fraud, gift cards could be manually distributed via text or email after each survey is verified." [30]	21.1%
	Not paying fraudulent respondents	Informing potential participants in the consent form that fraudulent participants will not be compensated. This helps researchers not waste money on fraudulent participants.		11.1%

XSL•FO
RenderX

Effectiveness ranking and technique		Description	Example	Freq.[a]
			"On the consent form, participants were also informed that "…we have put in place a number of safeguards to ensure that participants provide valid and accurate data for this study. If we have strong reason to believe your data are invalid, your responses will not be approved or paid and your data will be discarded." [31]	
No longer effective				
	IP address or Geolocation	Examining participant IP and geolocation to see if they match location requirements of study and screening for duplicate IP addresses.	"[Researchers used] other survey platform features to track IP addresses, geolocation, latitude and longitude, and participants' postal codes when they discovered that geographic markers or indicators did not match the participants' stated location of residence." [32]	65.6%
	(re) CAPTCHA	Including tests that can help to screen out bots by providing challenges that theoretically only humans can complete.	"Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) script was created and implemented into the Google Form." [33]	45.6%
	Timing checks	Checking for unusually fast or slow response times, which can indicate bots or satisficing.	"We noticed a large proportion of responses with improbably fast completion times (as well as those with particularly long completion times, eg, 4220 min)." [34]	45.6%
	Open-ended questions	Including open-ended questions and reviewing the responses. This can help assess attention as well as check for bots who may incoherently respond.	"Another indicator of data quality is suspicious responses to open-ended questions. For example, when given an open response box to report thoughts or ask questions at the end of the survey, responses written in all caps, one-word responses seemingly unrelated to the prompt, restatements of parts of the question, or nonsensical phrases." [35]	37.8%
	Honeypots	Incorporating questions into only the code of a survey, such that they are not visible to the human eye. These questions would only be answered by bots.	"We added a honeypot question as a second line of defense against bots. Honeypots are survey questions hidden from rendering on the screen using custom JavaScript code." [36]	14.4%

[a]Freq.: frequency of articles mentioning this technique.

Several techniques were deemed "somewhat effective," as they had notable benefits as well as limitations. For example, attention check questions help detect satisficers and some types of bots but are ineffective against fraudulent participants. Many studies noted suspicion when a large number of surveys were completed at once, as that can indicate their study has been "discovered" by bots or fraudulent respondents; thus, being watchful of a large number of responses is a useful technique

XSL•FO
RenderX

but is not sufficient for detecting all cases of threats to data integrity. Another technique, changing payment protocols (eg, intentionally not emphasizing participant payments in recruitment materials), was framed as a preventative measure rather than a definitive means of detecting fraud.

Finally, several techniques were deemed "no longer effective" by the authors: IP address and geolocation checks, reCAPTCHA, timing checks (ie, checking for unusually fast or unusually slow responses), open-ended questions, and honeypot questions (ie, questions not visible to the human eye but would be seen and answered by bots). Although many studies still report using these techniques, many authors also discussed how such methods can now be easily bypassed. For example, reCAPTCHA is not a challenge for most advanced bots, and certainly not for fraudulent participants. Likewise, proxy servers can help "fake" a local IP address. The invention of ChatGPT (OpenAI) and other artificial intelligence natural language processing chatbots makes short answer questions a less effective means of screening, as bots are often able to respond coherently to open-ended questions. Some bots can also be trained to complete surveys in a realistic timeframe and are also able to overlook honeypot questions. Therefore, although they are still somewhat useful for removing simpler bots and fraud attempts, these "no longer effective" techniques are unable to catch or detect more sophisticated attacks and should not be overly relied upon.

While discussing techniques to improve data integrity, many authors reflected on ethical considerations. For instance, as most "very effective" techniques require asking both personal and personally identifiable questions, additional procedures to protect participant privacy and rights may be necessary [37]. Another key ethical concern expressed by researchers was mistakenly excluding genuine participants. For example, fraud detection methods have the potential to introduce selection bias, such as when blocking responses from the same IP address deters residents of high-density housing developments [24]. Deterring genuine participants is also a major concern; for example, many techniques, such as requiring a video call at screening, can place additional burdens on participants and feel invasive [4]. Similarly, verification techniques that convey doubt to participants about their genuineness can compromise trust [38]. As a counter to these concerns, other authors discussed how techniques such as videoconferencing, when conducted with sensitivity, can foster and strengthen rapport and help researchers better understand their participant population [39]. Ultimately, thoughtful study design was encouraged, such as urging researchers to reflect on how to best balance their needs of (1) research integrity and quality; (2) feasibility and efficiency; and (3) safeguarding participants' rights, safety, and privacy [38]. Often, somewhat minor changes can help work toward this balance. For example, Singh and Sagar [37] encourage methods such as deidentifying data, using encryption processes or password-protected data storage, and using HIPAA (Health Insurance Portability and Accountability Act)-compliant online survey platforms. And Roehl and Harland [5] emphasized the importance of transparency during the consent process, so that participants are aware of what identifiable information will be requested from them and why.

Discussion

Principal Findings

Online research is expanding and holds great promise for innovative and impactful research. But as techniques to protect data integrity advance, so too do the methods of mendacious individuals providing false or unreliable responses for monetary gain. In this scoping review, we identified 90 articles published since 2020 that described methods for addressing online threats to data integrity [2-4,11,13,15,20-103]. We found that some of the most common techniques discussed were IP Address or Geolocation checks and reCAPTHCA. This is concerning, given that several articles detailed the reasons these techniques are no longer effective against sophisticated bots or fraudulent participants. Overall, these findings reveal a crucial area for improvement in handling threats to online data integrity. Yet our review also discovered new and innovative techniques for addressing threats to online data integrity. Specifically, we found that authenticating personal information, posing background-related questions, and cross-checking inconsistent answers were deemed very effective techniques for addressing contemporary threats.

Recommendations

While there is no one foolproof way for researchers to prevent participant fraud, it is clear from this review that the field has moved beyond reCAPTCHA as a sufficient technique for ensuring data integrity. Bots are advancing and fraudulent participants are becoming more sophisticated, making reCAPTCHA ill-equipped to handle the scope of the current problem. We recommend that researchers engaging in online data collection develop a robust strategy for ensuring data integrity early in the design of their research protocol. For such designs, we recommend that researchers use multiple techniques (rather than relying on the soundness of just one technique), use the techniques described in the articles reviewed here, and draw the most from techniques deemed to be "very effective." Although the constraints of timelines, person power, and budgets can render some techniques unfeasible, many "very effective" techniques are efficient and low cost; for example, researchers can think creatively to develop background-related questions that, even with internet searching, would only be readily known and (or) accurately answered by their population. When relying on survey service platforms for access to online samples, researchers should be critical of the techniques used by the platforms to guarantee the quality of their panels. We also suggest that techniques for ensuring data integrity should be critically considered by journal editors, journal reviewers, and grant reviewers when evaluating the rigor of study methods.

The ethical concerns discussed by the articles in this review highlight the responsibilities of researchers to continue focusing on participant rights and privacy. Techniques for ensuring data integrity (eg, personal questions to authenticate identity) should be balanced against these responsibilities. More broadly, the comfort of participants and their rapport with the study team should be considered. The relative weight of these considerations will vary depending on many factors. For instance, more complex study designs (eg, longitudinal studies that rely on

XSL•FO
RenderX

participant trust and investment for good retention) may require techniques that integrate authentication checks with rapport-building. The vulnerability of the sample and the sensitivity of the research topic must also be considered. Techniques for improving data integrity that increase participant burden, barriers, or privacy risk must be matched with greater participant accommodations—this could include greater compensation, clear explanations to participants during consent and enrollment about why certain questions are being asked, and enhanced data protection measures. Researchers attempting to exclude fraudulent participants should always be aware of their own biases and ensure that they are not excluding participants simply because they do not align with expected results.

Limitations and Future Directions

Our scoping review has some limitations. First, we only used one database for review (PubMed). This was done due to the large volume of articles on the subject and our focus on health research; however, it may have overlooked insights from other fields. Second, our search terms led to the studies reviewed primarily focusing on bots, fraudulent participants, and repeat participants, likely leading to an overestimation of these behaviors and an underestimation of the prevalence of participant satisficing and the tactics used to mitigate those behaviors. It is important to acknowledge that moving forward, some tactics that are currently in our "very effective" category may become less effective with the evolution of artificial intelligence, or as fraudulent participants become more familiar with current strategies. Going forward, more empirical studies should be conducted on research methods for addressing threats to data integrity, to quantitatively compare the effectiveness of various techniques used to address threats to data integrity in online research. Future work should also consider the participant perspective on these various techniques in order to improve their effectiveness and minimize negative consequences.

Conclusions

Threats to data integrity appear to be on the rise, particularly with online research, and numerous solutions and prevention strategies have been recommended. In order to aid researchers in developing robust online methodologies, this scoping review discusses the most common types of threats to data integrity, synthesizes the most common prevention methods, and discusses the ethical considerations raised about the techniques. Doubtless, new threats to data integrity will continue to emerge, and researchers should continue developing the most effective methods in response.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute and Food and Drug Administration's Center for Tobacco Products (CTP; grant U54CA287392). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Food and Drug Administration.

Data Availability

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

Authors' Contributions

IBS contributed to conceptualization (lead), data curation (lead), formal analysis (equal), investigation (lead), methodology (equal), and writing—original draft (lead). AKF was responsible for funding acquisition (equal), methodology (supporting), and writing—review and editing (equal). APT, JGP, and NJKB contributed to methodology (supporting) and writing—review and editing (equal). JD participated in formal analysis (equal) and writing—review and editing (equal). MER contributed to conceptualization (supporting), funding acquisition (equal), formal analysis (equal), investigation (supporting), methodology (supporting), supervision (lead), and writing—review and editing (lead).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study details described in the 90 manuscripts examined in this scoping review.
[DOCX File, 56 KB - ojphi_v17i1e70926_app1.docx]

Multimedia Appendix 2
Techniques for addressing threats to data integrity in online research as mentioned in the 90 manuscripts examined in this scoping review.
[DOCX File, 63 KB - ojphi_v17i1e70926_app2.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.

XSL•FO
RenderX

[DOCX File, 32 KB - ojphi_v17i1e70926_app3.docx]

References

1. Maheu C, Lemonde M, Mayo S, Galica J, Bally J. Moving research forward during COVID-19. Can Oncol Nurs J 2021;31(4):490-492. [Medline: 34786468]
2. Myers KJ, Jaffe T, Kanda DA, et al. Reaching the “Hard-to-Reach” sexual and gender diverse communities for population-based research in cancer prevention and control: methods for online survey data collection and management. Front Oncol 2022;12:841951. [doi: 10.3389/fonc.2022.841951] [Medline: 35756657]
3. Upadhyay UD, Jovel IJ, McCuaig KD, Cartwright AF. Using Google Ads to recruit and retain a cohort considering abortion in the United States. Contracept X 2020;2:100017. [doi: 10.1016/j.conx.2019.100017] [Medline: 32550532]
4. Pellicano E, Adams D, Crane L, et al. Letter to the Editor: a possible threat to data integrity for online qualitative autism research. Autism 2024 Mar;28(3):786-792. [doi: 10.1177/13623613231174543] [Medline: 37212144]
5. Roehl JM, Harland DJ. Imposter participants: overcoming methodological challenges related to balancing participant privacy with data quality when using online recruitment and data collection. Qual Rep 2022;27:2469-2485. [doi: 10.46743/2160-3715/2022.5475]
6. Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opin Q 1987;51(2):201. [doi: 10.1086/269029]
7. Suskie LA. Survey research: what works for the institutional researcher. institutional research information series no. 1. : North East Association for Institutional Research, C/O Larry Metzger, Ithaca College; 1988.
8. Bauermeister J, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data quality in web-based HIV/AIDS research: handling invalid and suspicious data. Field Methods 2012 Aug 1;24(3):272-291. [doi: 10.1177/1525822X12443097] [Medline: 23180978]
9. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to “fraudsters” in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133. [doi: 10.1111/jlme.12200] [Medline: 25846043]
10. Pullen Sansfaçon A, Gravel E, Gelly MA. Dealing with scam in online qualitative research: strategies and ethical considerations. Int J Qual Methods 2024 Jan;23. [doi: 10.1177/16094069231224610]
11. Lawrence PR, Osborne MC, Sharma D, Spratling R, Calamaro CJ. Methodological challenge: addressing bots in online research. J Pediatr Health Care 2023;37(3):328-332. [doi: 10.1016/j.pedhc.2022.12.006] [Medline: 36717299]
12. Al-Fannah NM. Making defeating captchas harder for bots 2017. arXiv. Preprint posted online on Apr 10, 2017 URL: https://doi.org/10.48550/arXiv.1704.02803 [accessed 2025-08-11]
13. Perkel JM. Mischief-making bots attacked my scientific survey. Nature New Biol 2020 Mar;579(7799):461-461. [doi: 10.1038/d41586-020-00768-0] [Medline: 32184487]
14. Storozuk A, Ashley M, Delage V, Maloney EA. Got bots? Practical recommendations to protect online survey data from bot attacks. Quant Method Psychol 2020;16(5):472-481. [doi: 10.20982/tqmp.16.5.p472]
15. Chandler J, Sisso I, Shapiro D. Participant carelessness and fraud: consequences for clinical research and potential solutions. J Abnorm Psychol 2020 Jan;129(1):49-55. [doi: 10.1037/abn0000479] [Medline: 31868387]
16. Devine EG, Pingitore AM, Margiotta KN, et al. Frequency of concealment, fabrication and falsification of study data by deceptive subjects. Contemp Clin Trials Commun 2021 Mar;21:100713. [doi: 10.1016/j.conctc.2021.100713] [Medline: 33604482]
17. Siegel JT, Navarro MA, Thomson AL. The impact of overtly listing eligibility requirements on MTurk: an investigation involving organ donation, recruitment scripts, and feelings of elevation. Soc Sci Med 2015 Oct;142:256-260. [doi: 10.1016/j.socscimed.2015.08.020] [Medline: 26322721]
18. Siegel JT, Navarro M. A conceptual replication examining the risk of overtly listing eligibility criteria on Amazon’s Mechanical Turk. J Applied Social Pyschol 2019 Apr;49(4):239-248. [doi: 10.1111/jasp.12580]
19. Sharpe Wessling K, Huber J, Netzer O. MTurk character misrepresentation: assessment and solutions. J Consum Res 2017 Jun 1;44(1):211-230. [doi: 10.1093/jcr/ucx053]
20. Gordon JH, Fujinaga-Gordon K, Sherwin C. Fraudulent online survey respondents may disproportionately threaten validity of research in small target populations. Health Expect 2024 Jun;27(3):e14099. [doi: 10.1111/hex.14099] [Medline: 38845165]
21. Hardesty JJ, Crespi E, Nian Q, et al. The vaping and patterns of e-cigarette use research study: protocol for a web-based cohort study. JMIR Res Protoc 2023 Mar 2;12:e38732. [doi: 10.2196/38732] [Medline: 36862467]
22. Davidson RD, Biddle K, Nassan M, Scammell TE, Zhou ES. The impact of narcolepsy on social relationships in young adults. J Clin Sleep Med 2022 Dec 1;18(12):2751-2761. [doi: 10.5664/jcsm.10212] [Medline: 35946418]
23. Tannahill HS, Blais RK. Using military screening questions to anonymously recruit post-9/11 era service members and veterans using online survey methods. Mil Med 2024 May 18;189(5-6):e1282-e1288. [doi: 10.1093/milmed/usad469] [Medline: 38140962]
24. Guest JL, Adam E, Lucas IL, et al. Methods for authenticating participants in fully web-based mobile app trials from the iReach Project: cross-sectional study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e28232. [doi: 10.2196/28232] [Medline: 34463631]

XSL•FO
RenderX

25. Webb MA, Tangney JP. Too good to be true: bots and bad data from Mechanical Turk. Perspect Psychol Sci 2024 Nov;19(6):887-890. [doi: 10.1177/17456916221120027] [Medline: 36343213]
26. Mizerek E, Wolf L, Moon MD. Identifying and mitigating fraud when using social media for research recruitment. J Emerg Nurs 2023 Jul;49(4):530-533. [doi: 10.1016/j.jen.2023.04.002] [Medline: 37393079]
27. Chen AT, Komi M, Bessler S, Mikles SP, Zhang Y. Integrating statistical and visual analytic methods for bot identification of health-related survey data. J Biomed Inform 2023 Aug;144:104439. [doi: 10.1016/j.jbi.2023.104439] [Medline: 37419375]
28. Hohn KL, Braswell AA, DeVita JM. Preventing and protecting against internet research fraud in anonymous web-based research: protocol for the development and implementation of an anonymous web-based data integrity plan. JMIR Res Protoc 2022 Sep 12;11(9):e38550. [doi: 10.2196/38550] [Medline: 36094806]
29. Bybee S, Cloyes K, Ellington L, Baucom B, Supiano K, Mooney K. Bots and nots: safeguarding online survey research with underrepresented and diverse populations. Psychol Sex 2022;13(4):901-911. [doi: 10.1080/19419899.2021.1936617] [Medline: 36439051]
30. Salem M, Pollack L, Zepeda A, Tebb KP. Utilization of online systems to promote youth participation in research: a methodological study. World J Methodol 2023 Sep 20;13(4):210-222. [doi: 10.5662/wjm.v13.i4.210] [Medline: 37771869]
31. Gratz KL, Tull MT, Richmond JR, Edmonds KA, Scamaldo KM, Rose JP. Thwarted belongingness and perceived burdensomeness explain the associations of COVID-19 social and economic consequences to suicide risk. Suicide Life Threat Behav 2020 Dec;50(6):1140-1148. [doi: 10.1111/sltb.12654] [Medline: 32589811]
32. Kumarasamy V, Goodfellow N, Ferron EM, Wright AL. Evaluating the problem of fraudulent participants in health care research: multimethod pilot study. JMIR Form Res 2024 Jun 4;8:e51530. [doi: 10.2196/51530] [Medline: 38833292]
33. Bisdas S, Topriceanu CC, Zakrzewska Z, et al. Artificial intelligence in medicine: a multinational multi-center survey on the medical and dental students' perception. Front Public Health 2021;9:795284. [doi: 10.3389/fpubh.2021.795284] [Medline: 35004598]
34. Burnette CB, Luzier JL, Bennett BL, et al. Concerns and recommendations for using Amazon MTurk for eating disorder research. Int J Eat Disord 2022 Feb;55(2):263-272. [doi: 10.1002/eat.23614] [Medline: 34562036]
35. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. PLoS One 2023;18(3):e0279720. [doi: 10.1371/journal.pone.0279720] [Medline: 36917576]
36. Bonett S, Lin W, Sexton Topper P, et al. Assessing and improving data integrity in web-based surveys: comparison of fraud detection systems in a COVID-19 study. JMIR Form Res 2024 Jan 12;8:e47091. [doi: 10.2196/47091] [Medline: 38214962]
37. Singh S, Sagar R. A critical look at online survey or questionnaire-based research studies during COVID-19. Asian J Psychiatr 2021 Nov;65:102850. [doi: 10.1016/j.ajp.2021.102850] [Medline: 34534919]
38. Wright M, Matheson J, Watson TM, Sproule B, Le Foll B, Brands B. Participant fraud in virtual qualitative substance use research: recommendations and considerations for detection and prevention based on a case study. Subst Use Misuse 2024;59(8):1261-1270. [doi: 10.1080/10826084.2024.2330892] [Medline: 38503716]
39. Wood NK, Bindler RJ. A videoconferencing verification method for enrollment of breastfeeding dyads to an online prospective mixed methods study during the COVID-19 pandemic. J Adv Nurs 2024 Jul;80(7):2970-2976. [doi: 10.1111/jan.15981] [Medline: 38012846]
40. Agley J, Xiao Y, Nolan R, Golzarri-Arroyo L. Quality control questions on Amazon's Mechanical Turk (MTurk): a randomized trial of impact on the USAUDIT, PHQ-9, and GAD-7. Behav Res Methods 2022 Apr;54(2):885-897. [doi: 10.3758/s13428-021-01665-8] [Medline: 34357539]
41. Bethel C, Rainbow JG, Dudding KM. Recruiting nurses via social media for survey studies. Nurs Res 2021;70(3):231-235. [doi: 10.1097/NNR.0000000000000482] [Medline: 33060416]
42. Bush J, Blackwell CW. Social media as a recruitment strategy with transgender-identified individuals: using an ethical lens to direct methodology. J Transcult Nurs 2022 Sep;33(5):603-614. [doi: 10.1177/10436596221101928] [Medline: 35699438]
43. Campbell CK, Ndukwe S, Dubé K, Sauceda JA, Saberi P. Overcoming challenges of online research: measures to ensure enrollment of eligible participants. JJ Acquir Immune Defic Syndr 2022;91(2):232-236. [doi: 10.1097/QAI.0000000000003035] [Medline: 36094490]
44. Davies MR, Monssen D, Sharpe H, et al. Management of fraudulent participants in online research: practical recommendations from a randomized controlled feasibility trial. Int J Eat Disord 2024 Jun;57(6):1311-1321 [FREE Full text] [doi: 10.1002/eat.24085] [Medline: 37921564]
45. Drysdale K, Wells N, Smith AKJ, Gunatillaka N, Sturgiss EA, Wark T. Beyond the challenge to research integrity: imposter participation in incentivised qualitative research and its impact on community engagement. Health Sociol Rev 2023 Sep 2;32(3):372-380. [doi: 10.1080/14461242.2023.2261433] [Medline: 37786312]
46. Dulin P, Mertz R, Edwards A, King D. Contrasting a mobile app with a conversational chatbot for reducing alcohol consumption: randomized controlled pilot trial. JMIR Form Res 2022;6(5):e33037. [doi: 10.2196/33037] [Medline: 35576569]
47. Dutra LM, Farrelly MC, Bradfield B, Ridenhour J, Guillory J. Modeling the Probability of Fraud in Social Media in a National Cannabis Survey: RTI Press; 2021.

XSL•FO
RenderX

48. Geiger G, Kiel L, Horiguchi M, et al. Latinas in medicine: evaluating and understanding the experience of Latinas in medical education: a cross sectional survey. BMC Med Educ 2024 Jan 3;24(1):4. [doi: 10.1186/s12909-023-04982-y]
49. Glazer JV, MacDonnell K, Frederick C, Ingersoll K, Ritterband LM. Liar! Liar! Identifying eligibility fraud by applicants in digital health research. Internet Interv 2021 Sep;25:100401. [doi: 10.1016/j.invent.2021.100401]
50. Godinho A, Schell C, Cunningham JA. Out damn bot, out: recruiting real people into substance use studies on the internet. Subst Abus 2020 Jan;41(1):3-5. [doi: 10.1080/08897077.2019.1691131]
51. Gonzalez JM, Grover K, Leblanc TW, Reeve BB. Did a bot eat your homework? An assessment of the potential impact of bad actors in online administration of preference surveys. PLoS ONE 2023 Oct 5;18(10):e0287766. [doi: 10.1371/journal.pone.0287766]
52. Gratz KL, Scamaldo KM, Vidaña AG, Richmond JR, Tull MT. Prospective interactive influence of financial strain and emotional nonacceptance on problematic alcohol use during the COVID-19 pandemic. Am J Drug Alcohol Abuse 2021 Jan 2;47(1):107-116. [doi: 10.1080/00952990.2020.1849248]
53. Griffin M, Martino RJ, LoSchiavo C, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant 2022 Aug;56(4):2841-2852. [doi: 10.1007/s11135-021-01252-1]
54. Guastaferro K, Abuchaibe V, McCormick KV, et al. Adapting a selective parent-focused child sexual abuse prevention curriculum for a universal audience: a pilot study. PLoS One 2024 May 16;19(5):e0302982. [doi: 10.1371/journal.pone.0302982]
55. Habib D, Jha N. AIM against survey fraud. JAMIA Open 2021 Oct 8;4(4):ooab099. [doi: 10.1093/jamiaopen/ooab099]
56. Hartman R, Moss AJ, Rabinowitz I, et al. Do you know the Wooly Bully? Testing era-based knowledge to verify participant age online. Behav Res Methods 2023 Oct;55(7):3313-3325. [doi: 10.3758/s13428-022-01944-y]
57. Hauser DJ, Moss AJ, Rosenzweig C, Jaffe SN, Robinson J, Litman L. Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. Behav Res Methods ;55(8):3953-3964. [doi: 10.3758/s13428-022-01999-x]
58. Ikegami K, Yoshimoto Y, Baba H, Sekoguchi S, Ando H, Ogami A. Study protocol and preliminary results of the impact of occupational health workers' activities on their health: Nationwide Prospective Internet-Based Survey. JMIR Form Res ;6(7):e35290. [doi: 10.2196/35290]
59. Ilagan MJ, Falk CF. Model-agnostic unsupervised detection of bots in a Likert-type questionnaire. Behav Res Methods ;56(5):5068-5085. [doi: 10.3758/s13428-023-02246-7]
60. Ilagan MJ, Falk CF. Supervised classes, unsupervised mixing proportions: detection of bots in a Likert-type questionnaire. Educ Psychol Meas 2023 Apr;83(2):217-239. [doi: 10.1177/00131644221104220]
61. Karuchit S, Thiengtham P, Tanpradech S, et al. A web-based, respondent-driven sampling survey among men who have sex with men (Kai Noi): description of methods and characteristics. JMIR Form Res 2024;8:e50812. [doi: 10.2196/50812]
62. Keith MG, McKay AS. Too anecdotal to be true? Mechanical Turk is not all bots and bad data: response to Webb and Tangney (2022). Perspect Psychol Sci 2024 Nov;19(6):900-907. [doi: 10.1177/17456916241234328]
63. Kim M, Lee H, Allison J. Challenges and lessons learned from a mobile health, web-based human papillomavirus intervention for female Korean American College Students: feasibility experimental study. JMIR Form Res 2020 Jan 29;4(1):e14111. [doi: 10.2196/14111]
64. Kolc KL, Tan YXK, Lo AZY, Shvetcov A, Mitchell PB, Perkes IE. Measuring psychiatric symptoms online: a systematic review of the use of inventories on Amazon Mechanical Turk (mTurk). J Psychiatr Res 2023 Jul;163:118-126. [doi: 10.1016/j.jpsychires.2023.05.027]
65. Lepage S, Conway A, Goodson N, Wicks P, Flight L, Devane D. Online randomised trials with children: a scoping review. PLoS One 2023 May 25;18(5):e0280965. [doi: 10.1371/journal.pone.0280965]
66. Levi R, Ridberg R, Akers M, Seligman H. Survey fraud and the integrity of web-based survey research. Am J Health Promot 2022 Jan;36(1):18-20. [doi: 10.1177/08901171211037531]
67. Loebenberg G, Oldham M, Brown J, et al. Bot or not? Detecting and managing participant deception when conducting digital research remotely: case study of a randomized controlled trial. J Med Internet Res 2023;25:e46523. [doi: 10.2196/46523]
68. Lorenzo-Luaces L, Howard J, Edinger A, et al. Sociodemographics and transdiagnostic mental health symptoms in SOCIAL (Studies of Online Cohorts for Internalizing Symptoms and Language) I and II: cross-sectional survey and botometer analysis. JMIR Form Res 2022;6(10):e39324. [doi: 10.2196/39324]
69. Mitchell JW, Chavanduka TMD, Sullivan S, Stephenson R. Recommendations from a descriptive evaluation to improve screening procedures for web-based studies with couples: cross-sectional study. JMIR Public Health Surveill 2020 May 12;6(2):e15079. [doi: 10.2196/15079]
70. Mournet AM, Kleiman EM. Internet-based mental health survey research: navigating internet bots on reddit. Cyberpsychol Behav Soc Netw 2023 Feb 1;26(2):73-79. [doi: 10.1089/cyber.2022.0173]
71. O'Donnell N, Satherley RM, Davey E, Bryan G. Fraudulent participants in qualitative child health research: identifying and reducing bot activity. Arch Dis Child 2023 May;108(5):415-416. [doi: 10.1136/archdischild-2022-325049] [Medline: 36669866]
72. Panesar P, Mayo SJ. "Taking out the trash": strategies for preventing and managing fraudulent data in web-surveys. Can Oncol Nurs J 2023;33(2):283-284. [Medline: 37152829]

XSL•FO
RenderX

73. Parks AM, Duffecy J, McCabe JE, et al. Lessons learned recruiting and retaining pregnant and postpartum individuals in digital trials: viewpoint. JMIR Pediatr Parent 2022;5(2):e35320. [doi: 10.2196/35320]
74. Pekarsky C, Skiffington J, Leijser LM, Slater D, Metcalfe A. Social media recruitment strategies to recruit pregnant women into a longitudinal observational cohort study: usability study. J Med Internet Res 2022;24(12):e40298. [doi: 10.2196/40298]
75. Pozzar R, Hammer MJ, Underhill-Blazey M, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020;22(10):e23021. [doi: 10.2196/23021]
76. Pratt-Chapman M, Moses J, Arem H. Strategies for the identification and prevention of survey fraud: data analysis of a web-based survey. JMIR Cancer 2021 Jul 16;7(3):e30730. [doi: 10.2196/30730] [Medline: 34269685]
77. Price M, Hidalgo JE, Kim JN, et al. The cyborg method: a method to identify fraudulent responses from crowdsourced data. Comput Human Behav 2024 Aug;157:108253. [doi: 10.1016/j.chb.2024.108253] [Medline: 38799787]
78. Reed ND, Bull S, Shrestha U, Sarche M, Kaufman CE. Combating fraudulent participation in Urban American Indian and Alaska Native Virtual Health Research: Protocol for Increasing Data Integrity in Online Research (PRIOR). JMIR Res Protoc 2024;13:e52281. [doi: 10.2196/52281]
79. Ridge D, Bullock L, Causer H, et al. 'Imposter participants' in online qualitative research, a new and increasing threat to data integrity? Health Expect 2023 Jun;26(3):941-944 [FREE Full text] [doi: 10.1111/hex.13724]
80. Rodriguez C, Oppenheimer DM. Creating a Bot-tleneck for malicious AI: psychological methods for bot detection. Behav Res Methods 2024 Sep;56(6):6258-6275. [doi: 10.3758/s13428-024-02357-9] [Medline: 38561551]
81. Rodriguez E, Peer K, Fruh V, et al. Digital global recruitment for women's health research: cross-sectional study. JMIR Form Res 2022;6(9):e39046. [doi: 10.2196/39046]
82. Roman ZJ, Brandt H, Miller JM. Automated bot detection using Bayesian latent class models in online surveys. Front Psychol 2022;13:789223. [doi: 10.3389/fpsyg.2022.789223]
83. Salinas MR. Are your participants real? Dealing with fraud in recruiting older adults online. West J Nurs Res 2023 Jan;45(1):93-99. [doi: 10.1177/01939459221098468] [Medline: 35587721]
84. Sefcik JS, Hathaway Z, DiMaria-Ghalili RA. When snowball sampling leads to an avalanche of fraudulent participants in qualitative research. Int J Older People Nurs 2023 Nov;18(6):e12572 [FREE Full text] [doi: 10.1111/opn.12572] [Medline: 37632269]
85. Sharma P, McPhail SM, Kularatna S, Senanayake S, Abell B. Navigating the challenges of imposter participants in online qualitative research: lessons learned from a paediatric health services study. BMC Health Serv Res 2020;24(1):724. [doi: 10.1186/s12913-024-11166-x]
86. Silva SSM, Meyer D, Jayawardana M. Detecting possible persons of interest in a physical activity program using step entries: Including a web ‐ based application for outlier detection and decision ‐ making. Biom J 2020 Mar;62(2):414-427 [FREE Full text] [doi: 10.1002/bimj.201900008]
87. Simone M, Cascalheira CJ, Pierce BG. A quasi-experimental study examining the efficacy of multimodal bot screening tools and recommendations to preserve data integrity in online psychological research. Am Psychol 2024;79(7):956-969. [doi: 10.1037/amp0001183]
88. Sosenko FL, Bramley G. Smartphone-based Respondent Driven Sampling (RDS): a methodological advance in surveying small or 'hard-to-reach' populations. PLoS One 2022;17(7):e0270673. [doi: 10.1371/journal.pone.0270673]
89. Stephenson R, Chavanduka TMD, Sullivan S, Mitchell JW. Correlates of successful enrollment of same-sex male couples into a web-based HIV prevention research study: cross-sectional study. JMIR Public Health Surveill 2020;6(1):e15078. [doi: 10.2196/15078]
90. Tran NK, Welles SL, Goldstein ND. Geolocation to identify online study-eligible gay, bisexual, and men who have sex with men in Philadelphia, Pennsylvania. Epidemiology 2023;34(4):462-466. [doi: 10.1097/EDE.0000000000001607]
91. Veldheer S, Whitehead-Zimmers M, Bordner C, et al. Participant preferences for the development of a digitally delivered gardening intervention to improve diet, physical activity, and cardiovascular health: cross-sectional study. JMIR Form Res 2023;7:e41498. [doi: 10.2196/41498]
92. Venugopal A, Marya A, Ludwig B, Alam MK. Satisficing and bots. Br Dent J 2022 Jan 28;232(2):68-68. [doi: 10.1038/s41415-022-3891-9]
93. Vu M, Huynh VN, Bednarczyk RA, et al. Experience and lessons learned from multi-modal internet-based recruitment of U.S. Vietnamese into research. PLoS One 2021;16(8):e0256074. [doi: 10.1371/journal.pone.0256074]
94. Walker LO, Murry N, Longoria KD. Improving data integrity and quality from online health surveys of women with infant children. Nurs Res 2023;72(5):386-391. [doi: 10.1097/NNR.0000000000000671]
95. Wang J, Calderon G, Hager ER, et al. Identifying and preventing fraudulent responses in online public health surveys: lessons learned during the COVID-19 pandemic. PLOS Glob Public Health 2023;3(8):e0001452. [doi: 10.1371/journal.pgph.0001452]
96. Wardropper CB, Dayer AA, Goebel MS, Martin VY. Conducting conservation social science surveys online. Conserv Biol 2021 Oct;35(5):1650-1658 [FREE Full text] [doi: 10.1111/cobi.13747]
97. Willis TA, Wright-Hughes A, Skinner C, et al. The detection and management of attempted fraud during an online randomised trial. Trials 2023;24(1):494. [doi: 10.1186/s13063-023-07517-4]

XSL•FO
RenderX

98. Woolfall K. Identifying and preventing fraudulent participation in qualitative research. Arch Dis Child 2023 Jun;108(6):421-422. [doi: 10.1136/archdischild-2023-325328]
99. Xu J, Tian G, He J, et al. The public's self-avoidance and other-reliance in the reporting of medical insurance fraud: a cross-sectional survey in China. Risk Manag Healthc Policy 2023;16:2869-2881. [doi: 10.2147/RMHP.S438854] [Medline: 38149180]
100. Xu J, Zhang T, Zhang H, et al. What influences the public's willingness to report health insurance fraud in familiar or unfamiliar healthcare settings? a cross-sectional study of the young and middle-aged people in China. BMC Public Health 2024;24(1):24. [doi: 10.1186/s12889-023-17581-9]
101. Young AM, Ballard AM, Cooper HLF. Novel recruitment methods for research among young adults in rural areas who use opioids: cookouts, coupons, and community-based staff. Public Health Rep 2020 Nov;135(6):746-755. [doi: 10.1177/0033354920954796]
102. Zhang H, Zhang T, Shi Q, et al. Who is more likely to report medical insurance fraud in the two scenarios of whether it results in a direct loss of individual benefit? A cross-sectional survey in China. Psychol Res Behav Manag 2022;15:2331-2341. [doi: 10.2147/PRBM.S375823]
103. Zuniga C, Ragosta S, Thompson TA. Recruiting foreign-born individuals who have sought an abortion in the United States: lessons from a feasibility study. Front Glob Womens Health 2024;4:1114820. [doi: 10.3389/fgwh.2023.1114820]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by E Mensah; submitted 06.01.25; peer-reviewed by H Maheshwari, O Oluwole, S Kath; revised version received 24.02.25; accepted 10.04.25; published 29.08.25.

Please cite as:
Strickland IB, Ferketich AK, Tackett AP, Patterson JG, Breitborde NJK, Davis J, Roberts M
Imposters, Bots, and Other Threats to Data Integrity in Online Research: Scoping Review of the Literature and Recommendations for Best Practices
Online J Public Health Inform 2025;17:e70926
URL: https://ojphi.jmir.org/2025/1/e70926
doi:10.2196/70926

© Isabella B Strickland, Amy K Ferketich, Alayna P Tackett, Joanne G Patterson, Nicholas J K Breitborde, Jade Davis, Megan Roberts. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 29.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Reporting of Fairness Metrics in Clinical Risk Prediction Models Used for Precision Health: Scoping Review

Lillian Rountree[1], BA; Yi-Ting Lin[1], ScM; Chuyu Liu[2], MS; Maxwell Salvatore[3], PhD; Andrew Admon[4], MD; Brahmajee Nallamothu[5], MD; Karandeep Singh[6], MD; Anirban Basu[7], PhD; Fan Bu[1], PhD; Bhramar Mukherjee[8], PhD

[1]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[2]Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing, MI, United States

[3]Departments of Biostatistics, Epidemiology, and Informatics and Genetics, University of Pennsylvania, Philadelphia, PA, United States

[4]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

[5]Institute for Healthcare Policy & Innovation, University of Michigan, Ann Arbor, MI, United States

[6]Department of Medicine, School of Medicine, University of California San Diego, San Diego, CA, United States

[7]The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, School of Pharmacy, University of Washington, Seattle, WA, United States

[8]Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT, United States

Corresponding Author:
Bhramar Mukherjee, PhD
Department of Biostatistics
Yale School of Public Health
Yale University
60 College St
New Haven, CT, 06510
United States
Phone: 1 2037378644
Email: bhramar.mukherjee@yale.edu

Abstract

Background: Clinical risk prediction models integrated into digitized health care informatics systems hold promise for personalized primary prevention and care, a core goal of precision health. Fairness metrics are important tools for evaluating potential disparities across sensitive features, such as sex and race or ethnicity, in the field of prediction modeling. However, fairness metric usage in clinical risk prediction models remains infrequent, sporadic, and rarely empirically evaluated.

Objective: We seek to assess the uptake of fairness metrics in clinical risk prediction modeling through an empirical evaluation of popular prediction models for 2 diseases, 1 chronic and 1 infectious disease.

Methods: We conducted a scoping literature review in November 2023 of recent high-impact publications on clinical risk prediction models for cardiovascular disease (CVD) and COVID-19 using Google Scholar.

Results: Our review resulted in a shortlist of 23 CVD-focused articles and 22 COVID-19 pandemic–focused articles. No articles evaluated fairness metrics. Of the CVD-focused articles, 26% used a sex-stratified model, and of those with race or ethnicity data, 92% had study populations that were more than 50% from 1 race or ethnicity. Of the COVID-19 models, 9% used a sex-stratified model, and of those that included race or ethnicity data, 50% had study populations that were more than 50% from 1 race or ethnicity. No articles for either disease stratified their models by race or ethnicity.

Conclusions: Our review shows that the use of fairness metrics for evaluating differences across sensitive features is rare, despite their ability to identify inequality and flag potential gaps in prevention and care. We also find that training data remain largely racially and ethnically homogeneous, demonstrating an urgent need for diversifying study cohorts and data collection. We propose an implementation framework to initiate change, calling for better connections between theory and practice when it comes to the adoption of fairness metrics for clinical risk prediction. We hypothesize that this integration will lead to a more equitable prediction world.

(Online J Public Health Inform 2025;17:e66598) doi:10.2196/66598

XSL•FO
RenderX

KEYWORDS

bias; cardiovascular disease; COVID-19; risk stratification; sensitive features; clinical risk prediction; equity

Introduction

Prediction models are increasingly prevalent in research and decision-making across a wide swath of fields, from finance to criminal justice to health care, and are a key building block in the practice and translation of precision health. Clinical prediction models, including updated versions of classical models like the Framingham Risk Score and the Gail Model [1,2], can leverage granular levels of multimodal data to optimize the implementation of precision health goals and enable the identification of high-risk individuals. However, potential statistical and historical biases ingrained in these models can impact their accuracy and ethical viability, particularly for populations at risk of discrimination due to sensitive features like race, ethnicity, age, or sex (Textbox 1) [3-6]. These risk prediction algorithms, designed to deliver individualized prevention strategies, are rarely evaluated across broad and coarse demographic categories, including the sensitive features mentioned above. Are these seemingly objective prediction models furthering existing inequities in precision health? Particularly as conversational and generative artificial intelligence tools are further integrated into the clinical prediction models that determine who is at high disease risk, who receives an intervention, or who receives prevention resources [6-8], and as the real-time implementation of these models in a hospital setting becomes more common [6], it is critical that these models are fair, providing accurate and nondiscriminatory predictions.

Textbox 1. Definitions of commonly used sensitive variables (also called sensitive features). Note that the definition of these variables is frequently left ambiguous in practice, raising the possibility for harmfully conflating the social and the biological.

Sensitive variables and their definition

Social variables

Variables are based on an individual's lived experience. Frequently self-reported.

- **Race**

  A social construct based on perceived physical differences that often acts as a proxy for various social and health consequences of racism in modeling [9]. Frequently paired with ethnicity, though they are not strictly equivalent. Includes no biological information.

- **Ethnicity**

  A social construct based on shared culture, language, geography, religion, and history [10]. Frequently paired with race, though they are not strictly equivalent. Includes no biological information.

- **Gender**

  A social construct that refers to an individual's self-presentation in society, informed by culture, psychology, and society [11]. Though distinct from sex, this variable often indirectly captures sex's impact.

Biological variables

Variables are based on an individual's biology.

- **Genetic ancestry**

  Genetic similarities between people due to common ancestors [10]. Distinct from race. Strictly biological.

- **Sex**

  The biological factors used to categorize individuals as male, female, or intersex [11]. Since 2016, it has been a required covariate in NIH-funded research. Though distinct from gender, this variable often indirectly captures gender's impact.

Algorithmic fairness is closely related to but theoretically distinct from algorithmic bias, another important consideration for assessing model performance. For further discussion of the subtle differences between these concepts, we refer the reader to the nuanced comparisons in [12,13]. We focus on algorithmic fairness in the current paper. Algorithmic fairness is concerned with designing algorithms and artificial intelligence models in a way that minimizes or mitigates bias and ensures fair treatment for all individuals or groups affected by the model [12,13]. Achieving algorithmic fairness requires careful consideration of the design, development, and deployment of models, including the selection of appropriate training data and the incorporation of fairness metrics to evaluate model performance on test data.

In recent years, several metrics have been introduced to evaluate the fairness of prediction models [14], alongside various coding toolboxes for their computation and implementation [15,16]. These metrics differ from common prediction metrics measuring discrimination, accuracy, and calibration, which measure overall model performance [12]. These fairness metrics typically aim to assess differences in model predictions ( ) for a binary decision outcome Y (eg, treatment or no treatment) for different values of a sensitive categorical variable S (eg, male or female sex), thus giving a numerical sense of how fair or unfair a model

XSL•FO
RenderX

may be. The common practice is to provide point estimates of a fairness metric; while fair inferential methods have been proposed [17], uncertainty quantification and interval estimation are not yet widely adopted in fairness research. Some of the most cited and used fairness metrics are listed and summarized in Table 1.

Though quite simple in terms of their definition, such metrics can shed light on otherwise unseen disparities in a model or dataset. For these metrics to produce meaningful results, a clear interpretation of the predictor and sensitive variables used is required in addition to how they affect the outcome of interest *Y*. For the sensitive features that fairness metrics usually seek to assess, it is not always simple to explain what they represent. Sensitive features often have multiple definitions and interpretations, particularly when it comes to what is biological versus what is socially constructed. This ambiguity is especially important for sex and race or ethnicity, 2 of the most used variables in learning algorithms. Sex is a biological variable, but it will likely capture effects from an individual's gender, information that might be the actual cause behind the sex variable's recorded influence on an outcome of interest. In the case of race and ethnicity, these variables are often self-reported, and thus their exclusive social nature is not always made clear, allowing for the inaccurate and harmful conclusion that differences observed with such race and ethnicity variables have a biological basis. Textbox 1 offers definitions of different measures related to these 2 specific sensitive variables and highlights how interpretations can be conflated and may not align with their intended use.

Table 1. Commonly used fairness metrics [18-21]. =Prediction or decision, Y=Observed data, and S=Sensitive feature, in the case of S being a multi-group variable where a comparison with a reference or privileged group is meaningful.  and Y are binary variables.

Metric	Equation	Definition
Equalized odds		Both groups should have equal true positive and false positive rates.
Equal opportunity		Both groups should have equal true positive rates. A relaxed version of equalized odds.
Predictive equality		The rate of false positives (negative events categorized as positives) should be independent of the sensitive feature.
False negative rate parity		The rate of false negatives (positive events categorized as negatives) should be independent of the sensitive feature.
Predictive parity		Model precision should be the same for both groups.
Demographic parity		The prediction or decision should be independent of the sensitive feature.

The risk of including sensitive variables only to misinterpret them in a way that furthers inequity has led to an ongoing debate about the place of these sensitive features in clinical risk prediction models. Exclusion of such sensitive features is 1 possible solution, mostly discussed in the context of self-reported race or ethnicity [22], though this kind of "fairness through unawareness" is not always successful and can in some cases further inequities [23]. The use of appropriate fairness metrics is another possible solution. Even when fairness metrics are used, however, the operationalization regarding the measurement of sensitive variables like race or ethnicity can differ from culture to culture, posing further challenges in assessing fairness. For example, while the United States predominantly gathers data on race rather than ethnicity, many European countries focus instead on ethnicity and nationality as indications of socially constructed differences [24]. The use of fairness metrics for sensitive variables that mire diverse definitions across geography and clinical studies remains a complex issue. Sociocultural adaptation of the definition and measurement of sensitive variables is needed for appropriate adoption of fairness. As our focus is on the specific contextual application of fairness metrics, we leave the nuances of this broader issue to many other excellent papers that explore the concept at length [3-5,9,12,18,19,24].

As prediction models are increasingly embedded into electronic health records to realize the goal of precision health [25,26], determining the intended use of sensitive features is crucial. How widely such metrics are used or reported in the clinical risk prediction literature is rarely objectively quantified. Thus, to gather empirical evidence to substantiate our viewpoint, we sought to examine the usage of fairness metrics in clinical risk prediction through a scoping review of recently published risk prediction models in high-impact journals for two diseases: cardiovascular disease ("CVD," a long-studied chronic and noncommunicable disease) and COVID-19 (a recently emerged infectious disease). Before looking at the data, we hypothesized that there would be little reporting of fairness metrics in CVD research, where many studies span years and began long before discussions of fairness metrics in the prediction literature, but that the emergence of the COVID-19 pandemic would pose an

XSL•FO
RenderX

opportunity to more frequently incorporate modern advances in fairness metrics into predicting disease outcomes.

Methods

A scoping literature review was conducted for each disease of interest. Our outcomes of interest differed slightly between CVD and COVID-19: the clinical risk prediction models for CVD focused on the fatal and nonfatal risks of CVD, while the models for the COVID-19 pandemic focused on the risk of mortality and severe disease. We did not use a classic systematic literature review approach, as our priority was to expediently capture the trend in the highest impact papers. We sought only high-impact papers to understand the prevailing, dominant practices of the field. Figure S1 in Multimedia Appendix 1 details the reproducible steps taken and search terms used for data collection and extraction.

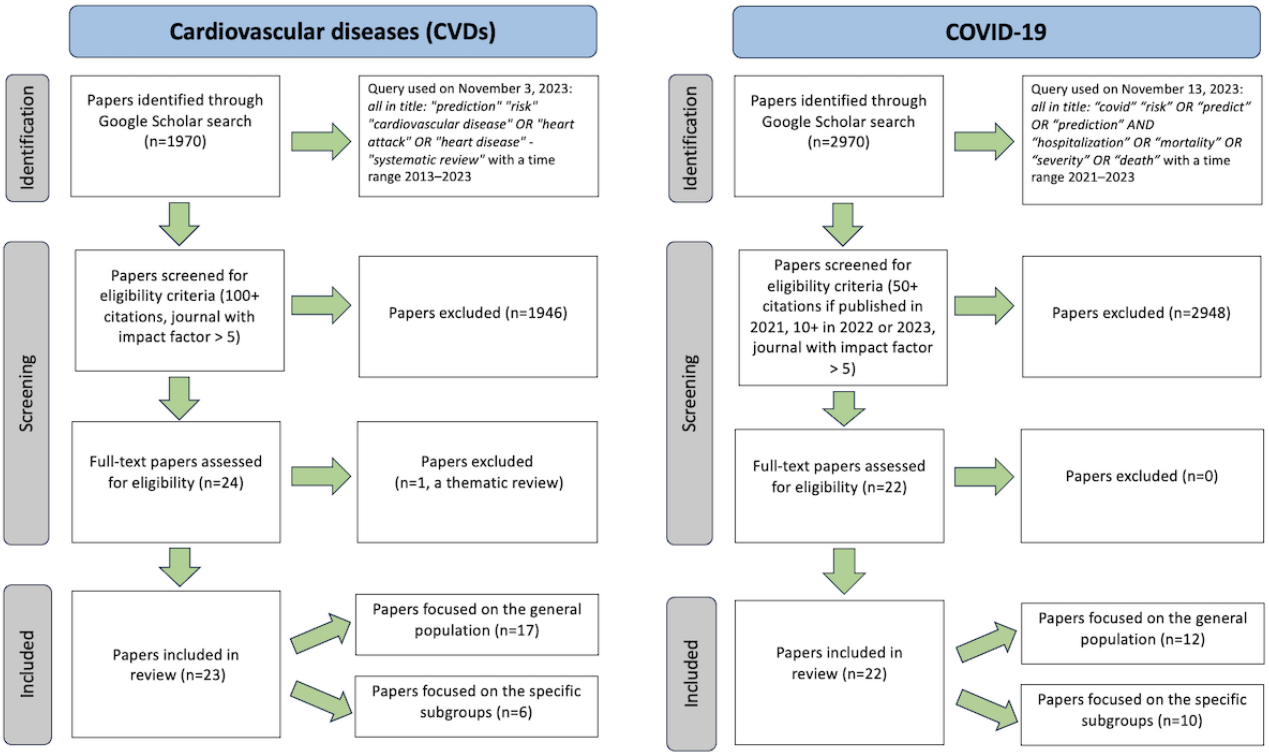
The criteria for highly impactful publications differed between CVD and COVID-19, reflecting the difference between a long-studied disease and an emerging area of study. CVD papers from the last 10 years (2013-2023) were reviewed and selected if they exceeded 100 citations and were from journals with impact factors exceeding 5, while COVID-19 papers from 2021 to 2023 were reviewed and selected if they exceeded 50 citations for 2021 papers or exceeded 10 citations for 2022 to 2023 papers, both from journals with impact factors over 5. Systematic reviews and meta-analysis papers were excluded from the results.

Results

Results From the Literature Review on CVD

The Google Scholar search query: allintitle: “prediction” “risk” “cardiovascular disease” OR “heart attack” OR “heart disease” OR “mortality” OR “death” -“systematic review” with a time range of 2013-2023 on November 3, 2023, returned 1970 results (Figure 1). For the purposes of our informal review, we reviewed the top 1000 returned by the Google Scholar algorithm. The yielded results were selected if they exceeded 100 citations and were from journals with impact factors exceeding 5. This provided a shortlist of 23 articles predicting the risk of a fatal or nonfatal CVD event that we then divided based on the target population (general population or a specific subpopulation). Tables S1 and S2 in Multimedia Appendix 2 and Figure S1 in Multimedia Appendix 1 of the supplementary material provide details.

Figure 1. PRISMA flow diagram of the literature search process.



Of the 17 CVD papers focusing on a general population that met the criteria of this review (Table S1 of Multimedia Appendix 2), none discussed fairness metrics. Of these 17 papers, 5 (29%) stratified their models by sex (ie, built different models for each sex), and 11 (65%) included sex as a covariate. Of these 17 papers, 9 (53%) included data on race or ethnicity (as many of the papers paired race and ethnicity together or used them interchangeably, we will refer to any discussion of either as race or ethnicity jointly, despite this being an imprecise practice). Only 5 of these 9 were multiracial or multiethnic (more than one racial or ethnic group identified), and 4 of these studies included race or ethnicity as a covariate. No study stratified its model by race or ethnicity (Figure 2). The 8 studies that did not include race or ethnicity data were based in the United States or Europe. Other sensitive features considered in the studies include body mass index, stratification by CVD mortality risk regions, and area-based measures of deprivation.

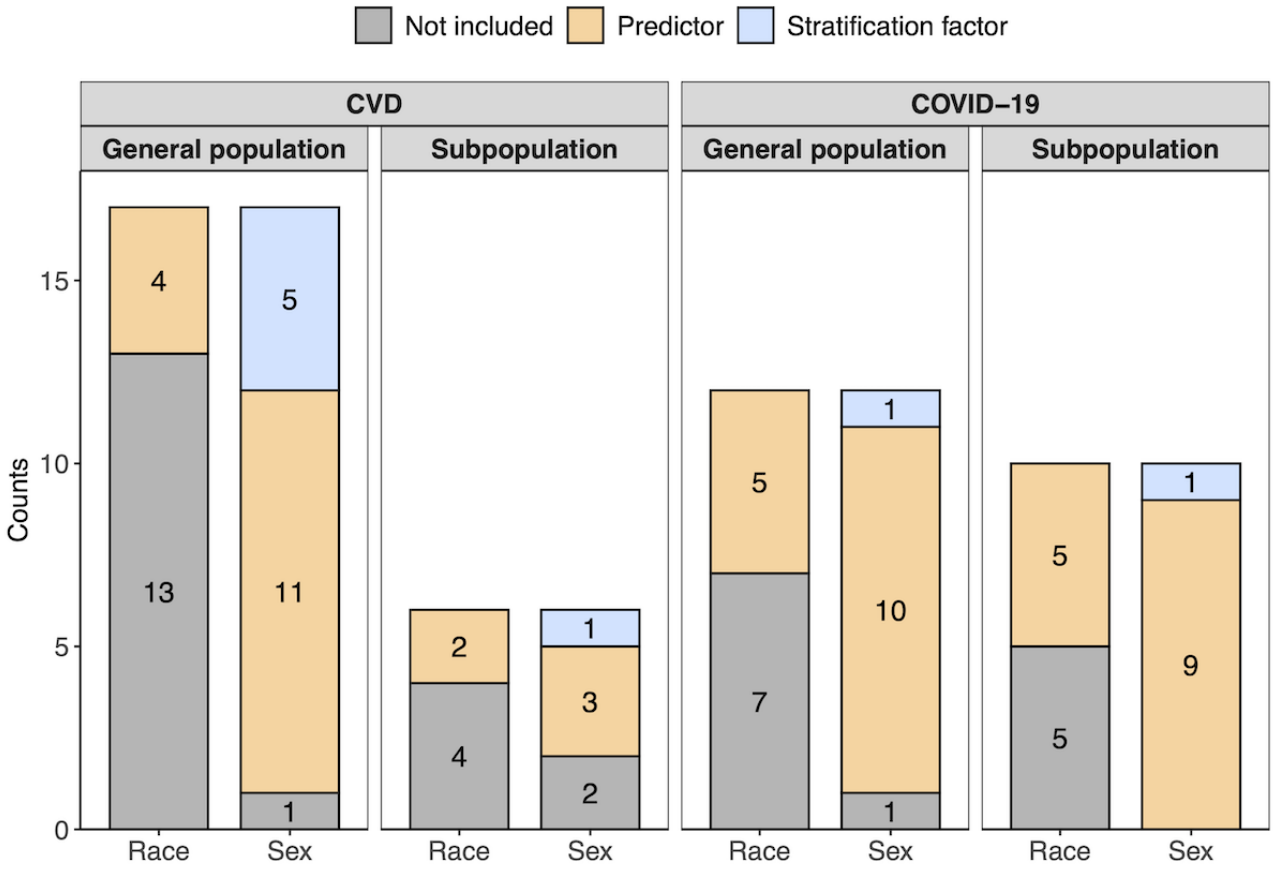
Similar results were observed in the 6 CVD papers that focused on specific subpopulations, ranging from those with chronic

XSL•FO
RenderX

conditions to those from specific ethnic or age groups (Table S2 in Multimedia Appendix 2). Of these 6 studies, 3 (50%) included sex as a covariate; only 1 stratified the prediction model by sex (Figure 2). All but one study included self-reported race or ethnicity data, yet only 3 (50%) were multiracial or multiethnic, and only 2 (33%) considered race or ethnicity as a covariate. Other sensitive features considered in the studies include geographic region, level of urbanization, and obesity.

Considering the general population and subpopulation papers together, 6 of the 23 (26%) CVD papers used a sex-stratified model, and of the 14 papers with race or ethnicity data, 13 (92%) had study populations where over 50% of participants were of a particular race or ethnicity.

Figure 2. Counts depicting how many of the reviewed articles include race or ethnicity and sex as either predictors or stratification factors in their clinical risk prediction models, separated by disease of interest and population focus of the article. For cardiovascular diseases (CVD), the outcome considered was the risk of cardiovascular disease, heart attack, or heart failure, whereas for the COVID-19 pandemic, it was hospitalization and death.



Results From the Literature Review on COVID-19

The Google Scholar search query: allintitle: “covid” “risk” “predict” OR “prediction” AND “hospitalization” OR “mortality” OR “severity” OR “death” with a time range of 2021-2023 on November 13, 2023, returned 2970 results. For the purposes of our informal review, we reviewed the top 1000 returned by the Google Scholar algorithm. These results were then selected if they exceeded 50 citations for 2021 papers and exceeded 10 citations for 2022 to 2023 papers, both from journals with impact factors over 5. This yielded a shortlist of 22 articles predicting the risk of COVID-19 hospitalization or death that we then divided based on the target population (general population or a specific subpopulation). Tables S3 and S4 in Multimedia Appendix 2 and Figure S1 in Multimedia Appendix 1 provide details.

Of the 12 COVID-19 papers (Table S3 in Multimedia Appendix 2) focusing on a general population that met the inclusion criteria, none mention fairness metrics. Of the 12 papers, 10 (83.3%) have sex as a covariate included in the prediction model; only one (8.3%) paper stratified participants by sex. These 5 of the 12 (41.7%) papers considered race or ethnicity, all of which included it as a covariate; none stratified by race or ethnicity (Figure 2). Of the remaining 7 papers, 3 cited a dearth of diverse data as the reason for lacking race or ethnicity information; 1 assumed the population to be entirely White; and the other 3 made no comments on race or ethnicity. Other sensitive covariates explored include patient-level socioeconomic index, geographical region, or hospital region.

Similar results were observed in the 10 COVID-19 papers that focused on specific subpopulations, ranging from those with pre-existing chronic diseases to age-specific groups like older adults and infants (Table S4 in Multimedia Appendix 2). All

XSL•FO
RenderX

10 papers included age as a risk factor. Similarly, all 10 considered sex in their models. One paper implemented stratification by sex, while all others (90%) used sex as a covariate. Of these 10 papers, 5 (50%) included race or ethnicity data, and all used it as a covariate. Though there was very little stratification by sex and none by race or ethnicity (Figure 2), stratification was undertaken for various other risk factors, such as the type of medication used and geographical location.

Considering the general population and subpopulation papers together, 2 of the 22 COVID-19 papers (9%) used a sex-stratified model, and of the 10 papers with race or ethnicity data, 5 (50%) had study populations where over 50% of participants were of a particular race or ethnicity.

Discussion

In our literature review, we found that assessing differences in model performance across sensitive features was somewhat common (eg, comparing areas under the receiver operating characteristic curves across sex or race), but discussing the implications of these differences for fairness and using appropriate fairness metrics were rare. There are several limitations of our approach. Our search was limited to highly cited papers in journals with high impact factors using the Google Scholar database and could have missed important articles; furthermore, we focused on only 2 diseases out of the many for which clinical risk prediction models are developed, restricting the generalizability of our findings to the broader class of clinical risk prediction models. Even though COVID-19 modeling began well after the field of fairness metrics was flourishing in quantitative research, in our review, their usage was as absent among COVID-19 studies as among CVD studies. It is our view that there are 2 primary reasons for the scarcity of fairness considerations in clinical risk prediction: a lack of diverse data and a lack of clarity on the best practice for using fairness metrics.

A lack of diverse data pervaded the studies in our review and was particularly prominent in terms of variation in race or ethnicity and geography. Though over half of the studies in the identified high-impact CVD-related papers reported on multiracial or multiethnic groups (as in, more than one racial or ethnic group reported) the data used in model development were still predominantly from 1 race or ethnicity. For example, in a study by Hippisley-Cox et al [27] with multiethnic data, though over 9 million patient records were used in the analysis, over 85% of subjects either self-reported their ethnic origin as White or did not report this information; other ethnic origins (like Indian, Black Caribbean, and other groups) made up less than 15% of patients. There was more racially or ethnically diverse data in the COVID-19–related papers: of the 10 papers with multiracial or multiethnic data, only half reported a study population of over 50% 1 race or ethnicity. The observed dominance across a single race or ethnicity implies that, for many studies, the lack of sufficient data to test-train models in each race-ethnic group separately makes fairness assessment challenging to begin with. In terms of geography, there was little diversity (Figure S2 in Multimedia Appendix 3) for either disease: the vast majority of reviewed studies were from the Global North.

Another limitation regarding the diversity of the data involved the definitions of race or ethnicity used in these studies. As mentioned in the Introduction, one important first step in using sensitive variables in risk prediction models is to clearly describe their intended use, such as whether they are designed to capture social attributes, biological features, or lived experiences. Of the 24 papers that had information on race or ethnicity, only 6 explicitly stated that the race or ethnicity variables were self-reported (25%); 2 papers directly described their study populations by genetic ancestry instead of by self-reported race or ethnicity (8.3%). The other 16 papers gave no contextualization to the way race or ethnicity data was measured, interpreted, and used, raising the risk of confusing the social as biological [23]. More exact details on the racial or ethnic makeup of the reviewed papers and how race or ethnicity was used and defined in each study are provided in Tables S1-S4 in Multimedia Appendix 2.

For some studies, however, such as those whose models were stratified by sex or study region, sufficient data did exist for computing and providing fairness metrics. Some of these studies did use discrimination and calibration metrics to assess the overall model fit across different subgroups (such as across race or ethnicity), though such analysis was not routinely conducted. Regardless, high discrimination and calibration alone do not guarantee model fairness, as aggregated data can easily mask inequalities [28,29]. From these studies, it appears that fairness analyses simply have not yet been adopted as a part of the model assessment routine.

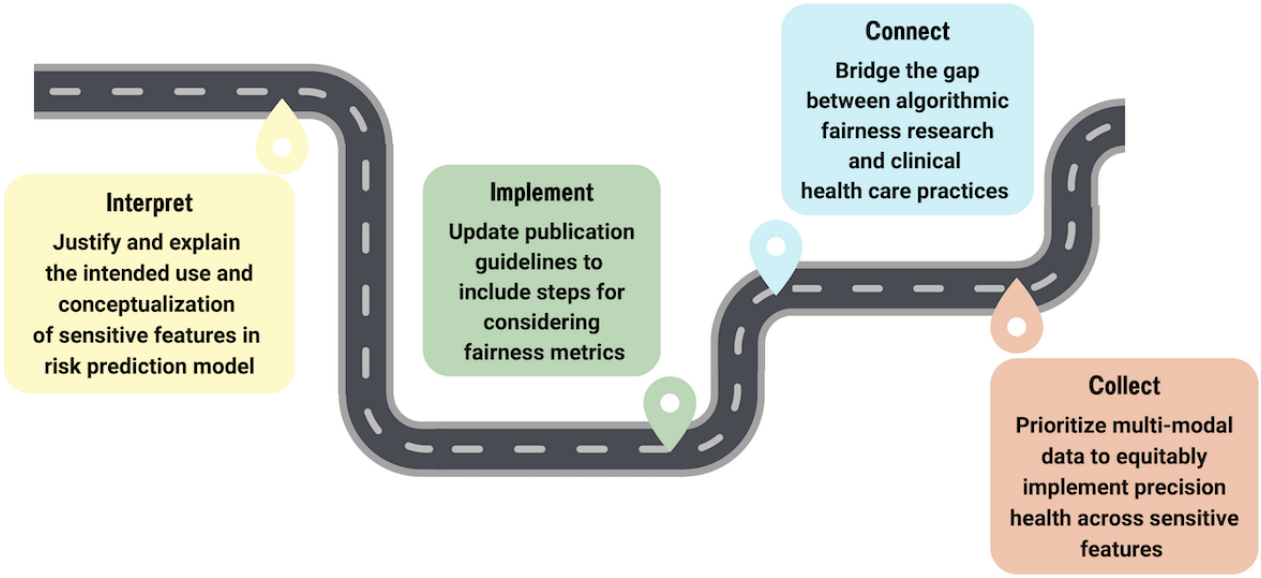
The lack of clarity as to how best to approach model fairness may have contributed to this practice. Even when the opportunity for the use of fairness metrics is obvious, the question of how to properly leverage them remains complex. There is no “one size fits all” method or universal fairness metric: instead, the specific context of the target outcome (whether it is a preventative intervention or a limited-resource treatment, for example) must inform how relevant concerns of fairness are, and what metrics can address the concerns [12]. Many fairness criteria are in fact mutually incompatible in practical settings (for example, demographic parity and equalized odds) [5], requiring a case-by-case decision on what kind of fairness (and thus fairness metric) will be most meaningful for the data and situation at hand. The limitations of many of the most common fairness metrics can also make their use inapplicable or unappealing for certain models. For example, most metrics are designed to assess only dichotomous outcomes and would require recalculation if a model’s predictions involved different cutoffs of the underlying continuous measures for different decisions. Yet, no critical number of developed fairness metrics will change practice; it is not an issue of lack of methods, but of implementation. Despite similar challenges, fairness methods have already been adopted in a variety of other predictive modeling fields, including criminal justice, finance, and computational linguistics [30-32].

XSL•FO
RenderX

One may pose the question if any clinical risk prediction modelers have used fairness metrics successfully but did not qualify for our scoping review due to its focus on only CVD and the COVID-19 pandemic and only highly cited papers in high-impact journals. To answer this question, we ran a high-level PubMed query of “fairness” AND “risk prediction” in all fields on January 6, 2025. This search returned 5 papers [33-37] on newly developed clinical risk prediction models for a variety of diseases, including preeclampsia, postpartum depression, and CVD, one of our diseases of interest. These papers made fairness metrics a core part of their model assessment and diagnostics but did not meet our criteria for inclusion in our review. For example, Liu et al [33] postpartum depression prediction model was first made to satisfy equal error rates between White and non-White individuals as the primary fairness metric; then debiasing techniques, including reweighing and fairness through blindness, were tested to improve the positive prediction rates between White and non-White individuals. These 5 studies acknowledged the limitations of fairness metrics and the impossibility of satisfying them all simultaneously and clearly defined their sensitive attributes. These new papers and others [38-40] can serve as helpful reference points for researchers as examples of attainable initial starts for fairness considerations when sensitive variables are measured.

Our own recommendations for how one can permeate and change practice in terms of adopting fairness as a core criterion and principle in clinical risk prediction are listed below and illustrated in the roadmap of Figure 3.

Figure 3. Strategies for increasing the adoption of fairness of clinical risk prediction models in precision health: Interpret, Implement, Connect and Collect (I2C2).



Roadmap and Recommendations for Changing Practice

We recommend the following 4 strategies for increasing the fairness of clinical risk prediction models across sensitive variables in precision health practice (I2C2).

- **Interpret**: In line with the NIH’s requirement of including (or justifying the exclusion of) a biological sex variable [41], papers should interpret, justify, and explain their intended use and conceptualization of sensitive features in risk prediction models.
- **Implement**: Influential guidelines like EQUATOR (Enhancing the Quality and Transparency of Health Research) TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guidelines [42] should include steps on considering algorithmic fairness as part of the implementation and application of clinical risk prediction models.
- **Connect**: The community of methods research in algorithmic fairness should ensure that the methods and tools developed are well broadcast to those in the community of practice in clinical health care, connecting theory to practice.
- **Collect**: The fields of clinical risk prediction and precision health should highly prioritize collecting inclusive, multimodel data across race, geographic region, and a variety of other sensitive or historically underrepresented features.

To understand the current barriers in the precision health practice community, we have developed a short questionnaire to help authors identify key challenges in implementing fairness metrics. This questionnaire will help clarify where resources should be most concentrated for this I2C2 roadmap. The questionnaire can be accessed here, and Table S5 in Multimedia Appendix 4 shows the summary questions. We hope by educating and enabling public health and clinical practitioners on the importance of assessing fairness metrics, we will create a more equitable prediction and precision prevention world for all.

XSL•FO
RenderX

All materials used for this review can be accessed via the GitHub of the University of Michigan Center for Precision Health Data Science.

Acknowledgments

This work was supported (grant DMS1712933) by the National Science Foundation and a MI-CARES grant (1UG3CA267907) from the National Cancer Institute. The funders had no role in the design of the study; collection, analysis, or interpretation of the data; writing of the report; or the decision to submit the manuscript for publication.

Data Availability

All materials used for this review can be accessed via the GitHub of the University of Michigan Center for Precision Health Data Science.

Authors' Contributions

Conceptualization and Investigation: LR, YTL, and CL.
Supervision: MS and BM.
Funding acquisition: BM.
Writing – original draft: YTL and LR.
Visualization: YTL.
Writing – review & editing: All authors.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 108 KB - ojphi_v17i1e66598_app1.docx]

Multimedia Appendix 2
Articles selected for review. Includes Tables S1-S4.
[PDF File (Adobe PDF File), 253 KB - ojphi_v17i1e66598_app2.pdf]

Multimedia Appendix 3
Counts depicting the geographic origin of data used in the reviewed articles, separated by disease of interest and population focus of the article. Studies with study regions covering multiple continents are double counted.
[PDF File (Adobe PDF File), 5 KB - ojphi_v17i1e66598_app3.pdf]

Multimedia Appendix 4
The ten questions included in the questionnaire. All but questions 9 and 10 are multiple choice, with the option to elaborate in a free response. Supplementary Table 5.
[DOCX File , 15 KB - ojphi_v17i1e66598_app4.docx]

References

1. Framingham Heart Study FHS. Cardiovascular disease (10-year risk). Framingham Heart Study. URL: https://www.framinghamheartstudy.org/fhs-risk-functions/cardiovascular-disease-10-year-risk/ [accessed 2025-02-20]
2. Breast cancer risk assessment tool: online calculator (the Gail Model). National Cancer Institute. URL: https://bcrisktool.cancer.gov/ [accessed 2024-11-19]
3. Bærøe K, Gundersen T, Henden E, Rommetveit K. Can medical algorithms be fair? Three ethical quandaries and one dilemma. BMJ Health & Care Informatics 2022 Apr 01;29(1):e100445 [FREE Full text] [doi: 10.1136/bmjhci-2021-100445] [Medline: 35396245]
4. Mitchell S, Potash E, Barocas S, D'Amour A, Lum K. Algorithmic fairness: choices, assumptions, and definitions. Annu. Rev. Stat. Appl 2021;8(1):141-163. [doi: 10.1146/annurev-statistics-042720-125902]
5. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput. Surv 2021;54(6):1-35 [FREE Full text] [doi: 10.1145/3457607]

XSL•FO
RenderX

6. Wiens J, Spector-Bagdady K, Mukherjee B. Toward realizing the promise of AI in precision health across the spectrum of care. Annual Review of Genomics and Human Genetics 2024;25(1):141-159 [FREE Full text] [doi: 10.1146/annurev-genom-010323-010230] [Medline: 38724019]
7. Biden JR. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Federal Register. URL: https://www.federalregister.gov/d/2023-24283 [accessed 2023-10-30]
8. Coiera E, Fraile-Navarro D. AI as an Ecosystem — Ensuring Generative AI Is Safe and Effective. NEJM AI 2024 Aug 22;1(9):AIp2400611. [doi: 10.1056/AIp2400611]
9. Paulus JK, Kent DM. Race and Ethnicity: A Part of the Equation for Personalized Clinical Decision Making? Circulation: Cardiovascular Quality and Outcomes 2017 Jul;10(7):e003823. [doi: 10.1161/circoutcomes.117.003823]
10. Lu C, Ahmed R, Lamri A, Anand SS. Use of race, ethnicity, and ancestry data in health research. PLOS Glob Public Health 2022;2(9):e0001060 [FREE Full text] [doi: 10.1371/journal.pgph.0001060] [Medline: 36962630]
11. Tannenbaum C, Ellis RP, Eyssel F, Zou J, Schiebinger L. Sex and gender analysis improves science and engineering. Nature 2019 Nov;575(7781):137-146. [doi: 10.1038/s41586-019-1657-6] [Medline: 31695204]
12. Paulus JK, Kent DM. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med 2020;3(1):99. [doi: 10.1038/s41746-020-0304-9] [Medline: 32821854]
13. Fletcher RR, Nakeshimana A, Olubeko O. Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. Front Artif Intell 2020;3:561802 [FREE Full text] [doi: 10.3389/frai.2020.561802] [Medline: 33981989]
14. Parbhoo S, Wawira Gichoya J, Celi LA, de la Hoz MA, for MIT Critical Data. Operationalising fairness in medical algorithms. BMJ Health Care Inform 2022 Jun;29(1):e100617 [FREE Full text] [doi: 10.1136/bmjhci-2022-100617] [Medline: 35688512]
15. Adebayo JA. Massachusetts institute of technology. FairML: ToolBox for diagnosing bias in predictive modeling. 2016. URL: http://hdl.handle.net/1721.1/108212 [accessed 2025-02-20]
16. Kozodoi N, Varga TV. Algorithmic Fairness in R. Nikita Kozodoi. 2020. URL: https://kozodoi.me/r/fairness/packages/2020/05/01/fairness-tutorial.html [accessed 2024-05-01]
17. Nabi R, Shpitser I. Fair inference on outcomes. In: AAAI. 2018 Presented at: Proceedings of the AAAI Conference on Artificial Intelligence; 2018; New Orleans, Louisiana, USA. [doi: 10.1609/aaai.v32i1.11553]
18. 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 [FREE Full text] [doi: 10.1038/s41598-022-07939-1] [Medline: 35273279]
19. Verma S, Rubin J. Fairness Definitions Explained. In: Association for Computing Machinery. 2018 Presented at: FairWare '18: Proceedings of the International Workshop on Software Fairness; 2018 May 29; Gothenburg, Sweden p. 1-7. [doi: 10.1145/3194770.3194776]
20. Chouldechova A. Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data 2017;5(2):153-163. [doi: 10.1089/big.2016.0047] [Medline: 28632438]
21. Hardt M, Price E, Srebro N. Equality of Opportunity in Supervised Learning. arXiv. 2016. URL: https://arxiv.org/abs/1610.02413 [accessed 2023-10-07]
22. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. N Engl J Med 2020 Aug 27;383(9):874-882 [FREE Full text] [doi: 10.1056/NEJMms2004740] [Medline: 32853499]
23. Basu A. Mitigating Clinical Algorithmic Discrimination. JAMA Health Forum 2024 Dec 06;5(12):e244190 [FREE Full text] [doi: 10.1001/jamahealthforum.2024.4190] [Medline: 39671202]
24. Jaime S, Kern C. Ethnic classifications in algorithmic fairness: concepts, measures, and implications in practice. 2024 Presented at: FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency; 2024 June 6; Rio de Janeiro, Brazil. [doi: 10.1145/3630106.3658902]
25. Sharma V, Ali I, van der Veer S, Martin G, Ainsworth J, Augustine T. Adoption of clinical risk prediction tools is limited by a lack of integration with electronic health records. BMJ Health Care Inform 2021 Feb;28(1):e100253 [FREE Full text] [doi: 10.1136/bmjhci-2020-100253] [Medline: 33608259]
26. Lee TC, Shah NU, Haack A, Baxter SL. Clinical Implementation of Predictive Models Embedded within Electronic Health Record Systems: A Systematic Review. Informatics (MDPI) 2020 Sep;7(3):25 [FREE Full text] [doi: 10.3390/informatics7030025] [Medline: 33274178]
27. Hippisley-Cox J, Coupland CAC, Mehta N, Keogh RH, Diaz-Ordaz K, Khunti K, et al. Risk prediction of covid-19 related death and hospital admission in adults after covid-19 vaccination: national prospective cohort study. BMJ 2021 Sep 17;374:n2244 [FREE Full text] [doi: 10.1136/bmj.n2244]
28. Loi M, Heitz C. Is calibration a fairness requirement? An argument from the point of view of moral philosophy and decision theory. 2022 Presented at: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency; 2022 June 24; Seoul, Republic of Korea. [doi: 10.1145/3531146.3533245]
29. Kauh TJ, Read JG, Scheitler AJ. The Critical Role of Racial/Ethnic Data Disaggregation for Health Equity. Popul Res Policy Rev 2021;40(1):1-7 [FREE Full text] [doi: 10.1007/s11113-020-09631-6] [Medline: 33437108]
30. Wang C, Han B, Patel B, Rudin C. In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction. J Quant Criminol 2022 Mar 28;39(2):519-581. [doi: 10.1007/s10940-022-09545-w]

XSL•FO
RenderX

31. Te YF, Wieland M, Frey M, Grabner H. Mitigating discriminatory biases in success prediction models for venture capitals. 2023 Presented at: 10th IEEE Swiss Conference on Data Science (SDS); 2023 June 23; Zurich, Switzerland. [doi: 10.1109/sds57534.2023.00011]
32. Schwertmann L, Kannan Ravi MP, de Melo G. Model-Agnostic Bias Measurement in Link Prediction. In: Findings of the Association for Computational Linguistics: EACL 2023. Croatia: Association for Computational Linguistics; 2023 Presented at: Association for Computational Linguistics: EACL 2023; 2023; Dubrovnik, Croatia URL: https://aclanthology.org/2023.findings-eacl.121 [doi: 10.18653/v1/2023.findings-eacl.121]
33. Liu Y, Joly R, Reading Turchioe M, Benda N, Hermann A, Beecy A, et al. Preparing for the bedside-optimizing a postpartum depression risk prediction model for clinical implementation in a health system. J Am Med Inform Assoc 2024;31(6):1258-1267 [FREE Full text] [doi: 10.1093/jamia/ocae056] [Medline: 38531676]
34. Atehortúa A, Gkontra P, Camacho M, Diaz O, Bulgheroni M, Simonetti V, et al. Cardiometabolic risk estimation using exposome data and machine learning. Int J Med Inform 2023;179:105209 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105209] [Medline: 37729839]
35. Topaz M, Davoudi A, Evans L, Sridharan S, Song J, Chae S, et al. Building a time-series model to predict hospitalization risks in home health care: insights into development, accuracy, and fairness. J Am Med Dir Assoc 2025;26(2):105417. [doi: 10.1016/j.jamda.2024.105417] [Medline: 39689864]
36. Li C, Jiang X, Zhang K. A transformer-based deep learning approach for fairly predicting post-liver transplant risk factors. J Biomed Inform 2024;149:104545 [FREE Full text] [doi: 10.1016/j.jbi.2023.104545] [Medline: 37992791]
37. Lin YC, Mallia D, Clark-Sevilla AO, Catto A, Leshchenko A, Yan Q, et al. A comprehensive and bias-free machine learning approach for risk prediction of preeclampsia with severe features in a nulliparous study cohort. BMC Pregnancy Childbirth 2024 Dec 24;24(1):853 [FREE Full text] [doi: 10.1186/s12884-024-06988-w] [Medline: 39716098]
38. Kartoun U, Khurshid S, Kwon BC, Patel AP, Batra P, Philippakis A, et al. Prediction performance and fairness heterogeneity in cardiovascular risk models. Sci Rep 2022 Jul 22;12(1):12542 [FREE Full text] [doi: 10.1038/s41598-022-16615-3] [Medline: 35869152]
39. Foryciarz A, Pfohl SR, Patel B, Shah N. Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation. BMJ Health Care Inform 2022;29(1):e100460 [FREE Full text] [doi: 10.1136/bmjhci-2021-100460] [Medline: 35396247]
40. Wastvedt S, Huling JD, Wolfson J. An intersectional framework for counterfactual fairness in risk prediction. Biostatistics 2024;25(3):702-717. [doi: 10.1093/biostatistics/kxad021] [Medline: 37660301]
41. Arnegard ME, Whitten LA, Hunter C, Clayton JA. Sex as a biological variable: a 5-year progress report and call to action. J Womens Health (Larchmt) 2020;29(6):858-864 [FREE Full text] [doi: 10.1089/jwh.2019.8247] [Medline: 31971851]
42. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med 2015 May;162(10):735-736. [doi: 10.7326/l15-5093-2]

Abbreviations

CVD: cardiovascular disease
EQUATOR: Enhancing the Quality and Transparency of Health Research
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

Edited by E Mensah; submitted 17.09.24; peer-reviewed by D Smolyak, C Kern; comments to author 12.12.24; revised version received 03.02.25; accepted 17.02.25; published 19.03.25.

Please cite as:
Rountree L, Lin YT, Liu C, Salvatore M, Admon A, Nallamothu B, Singh K, Basu A, Bu F, Mukherjee B
Reporting of Fairness Metrics in Clinical Risk Prediction Models Used for Precision Health: Scoping Review
Online J Public Health Inform 2025;17:e66598
URL: https://ojphi.jmir.org/2025/1/e66598
doi:10.2196/66598
PMID:39962044

©Lillian Rountree, Yi-Ting Lin, Chuyu Liu, Maxwell Salvatore, Andrew Admon, Brahmajee Nallamothu, Karandeep Singh, Anirban Basu, Fan Bu, Bhramar Mukherjee. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Accuracy of Nutrition-Related Awareness Messages on Twitter (Rebranded as X) by the Nutrition Awareness Providers in the Kingdom of Saudi Arabia: Validity Content Analysis

Duaa Alammari[1,2], MAS, DrPH; Iman A Bindayel[3], PhD; Noura Althukair[3], MSc; Noura AlRomi[3], BSc; Khalid Aldubayan[3], PhD; Najla Khateeb[4], PhD; Ahmed Alabdrabalnabi[5], DrPH; Banan Banamah, MPH

[1]Department of Health Systems Management, College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Prince Mutib Ibn Abdullah ibn Abdulaziz Rd, Riyadh, Saudi Arabia

[2]King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

[3]Department of Community Health and Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

[4]College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences, Al Ahsa, Saudi Arabia

[5]Department of Public Health, College of Health Sciences, Saudi Electronic University, Dammam, Saudi Arabia

Corresponding Author:
Duaa Alammari, MAS, DrPH
Department of Health Systems Management, College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Prince Mutib Ibn Abdullah ibn Abdulaziz Rd, Riyadh, Saudi Arabia

Related Article:

Correction of: https://ojphi.jmir.org/2025/1/e68128

Abstract

(Online J Public Health Inform 2025;17:e85577) doi:10.2196/85577

In "Accuracy of Nutrition-Related Awareness Messages on Twitter (Rebranded as X) by the Nutrition Awareness Providers in the Kingdom of Saudi Arabia: Validity Content Analysis" [1], the authors noted one error.

Author AA's affiliation was originally published as:

> *Department of Clinical Nutrition, College of Health Sciences, Saudi Electronic University, Dammam, Saudi Arabia*

This has now been corrected to:

> *Department of Public Health, College of Health Sciences, Saudi Electronic University, Dammam, Saudi Arabia*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Alammari D, Bindayel IA, Althukair N, et al. Accuracy of nutrition-related awareness messages on Twitter (rebranded as X) by the nutrition awareness providers in the Kingdom of Saudi Arabia: validity content analysis. Online J Public Health Inform 2025 Sep 26;17:e68128. [doi: 10.2196/68128] [Medline: 41004639]

Submitted 09.10.25; this is a non–peer-reviewed article;accepted 09.10.25; published 23.10.25.

Please cite as:
Alammari D, Bindayel IA, Althukair N, AlRomi N, Aldubayan K, Khateeb N, Alabdrabalnabi A, Banamah B
Correction: Accuracy of Nutrition-Related Awareness Messages on Twitter (Rebranded as X) by the Nutrition Awareness Providers in the Kingdom of Saudi Arabia: Validity Content Analysis
Online J Public Health Inform 2025;17:e85577
URL: https://ojphi.jmir.org/2025/1/e85577
doi:10.2196/85577

XSL•FO
RenderX

© Duaa Alammari, Iman A Bindayel, Noura Althukair, Noura AlRomi, Khalid Aldubayan, Najla Khateeb, Ahmed Alabdrabalnabi, Banan Banamah. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 23.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Bayesian Network Analysis of Intervention-Induced Physical Activity Behavior Change: Comparative Modeling Study Across Age, Education, and Activity Impairment Subgroups

Simone Catharina Maria Wilhelmina Tummers[1], MSc; Arjen Hommersom[2,3], MSc, PhD; Lilian Lechner[1], MSc, PhD; Roger Bemelmans[4], MSc, PhD; Catherine Adriana Wilhelmina Bolman[1], MSc, PhD

[1]Department of Health Psychology, Faculty of Psychology, Open University of the Netherlands, Valkenburgerweg 177, Heerlen, The Netherlands

[2]Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands

[3]Department of Computer Science, Faculty of Science, Open University of the Netherlands, Heerlen, The Netherlands

[4]Lectorate Data Intelligence, Zuyd University of Applied Sciences, Heerlen, The Netherlands

Corresponding Author:
Simone Catharina Maria Wilhelmina Tummers, MSc
Department of Health Psychology, Faculty of Psychology, Open University of the Netherlands, Valkenburgerweg 177, Heerlen, The Netherlands

Abstract

Background: Tailoring intervention content, such as those designed to improve physical activity (PA) behavior, can enhance effectiveness. Previous Bayesian network research showed that it might be relevant to tailor PA interventions based on demographic factors such as gender, revealing differences in determinants' roles between subpopulations. In order to optimize tailoring, one needs to understand the differences between subpopulations based on different characteristics. Building on this, this study examines age, education level, and PA impairment as key moderators, as these factors might influence PA engagement and intervention responsiveness. Older adults, for example, rely more on habitual behavior, lower-educated individuals may face challenges due to lower health literacy and socioeconomic inequalities, and individuals with PA impairment, defined as functional impairments restricting PA, may face unique barriers to PA. Understanding differences based on these factors is crucial for optimizing interventions and ensuring effectiveness across diverse populations.

Objective: This study investigates, by means of Bayesian networks, differences in PA intervention mechanisms of subpopulations based on age, education level, and PA impairment.

Methods: Subpopulation-specific subsets from an integrated dataset of 5 studies are analyzed, including demographics, experimental group assignment, and PA and sociocognitive measures at baseline, short term, and long term. The relevant subpopulations are defined based on age, education level, and PA impairment. For each subpopulation, a stable Bayesian network is estimated based on the corresponding subset of data by applying a bootstrap procedure and according to a confidence threshold, relevant paths of the model are visualized in order to find indications regarding subpopulation-specific intervention mechanisms.

Results: A comparison of subpopulation-specific models unveils similarities and differences with respect to determinants' roles in PA behavior change induced by interventions. Similar structures of determinants affect short-term PA, ultimately causing effects in the long term, where intention and habit are directly related to PA for most subpopulations. With respect to age-based differences, the interventions influence PA less via attitude cons and planning for older than younger people. Looking at the level of education, planning and intrinsic motivation are less influential for low-educated participants compared with high- or medium-educated participants, whereas more influence takes place through attitude pros for this low-educated group with respect to maintaining effects in the long term. Looking at PA impairments, apart from the findings that attitude pros and planning are more relevant in the pathway of change for people without impairment, a more interesting insight is that fewer determinants are directly influenced by the intervention within the group with PA impairment.

Conclusions: Intervention mechanisms in specific demographic groups have been rarely studied so far. Initial interpretations from the derived subpopulation models in this study unveil subpopulation-specific patterns of behavioral change, which enable better tailoring of intervention content to characteristics of the target population in order to induce or enhance effects.

(Online J Public Health Inform 2025;17:e57977) doi:10.2196/57977

XSL•FO
RenderX

KEYWORDS

physical activity; physical exercise; subpopulation differences; behavior change; age; education; physical activity impairment; Bayesian network; eHealth; e-health intervention; digital health; digital technology; digital intervention; multiple studies; demographics

Introduction

Despite known health benefits of physical activity (PA), adherence to the PA guidelines of the World Health Organization is low across different subpopulations [1-7]. E-health interventions have been designed to improve PA behavior of different target subpopulations, for example, Dutch-speaking adults older than 50 years [8], independently living, Dutch-speaking adults older than 64 years experiencing PA impairment [9], or (former) adult patients with prostate and colorectal cancer [10]. Such behavioral change intervention programs are designed based on behavioral change theories such as the theory of planned behavior and self-determination theory [11,12]. Based on these theories, behavior change strategies are incorporated into the programs in order to influence psycho-social determinants of PA behavior [13,14]. Determinants are factors expected to be directly or indirectly correlated with PA behavior, and it is presumed that changes in behavior are reached by influencing these determinants [8,9,13,15-18]. Tailoring of interventions based on an individual participant's characteristics, resulting in personalized intervention content, has proven effective [19-21]. It is known that different subgroups, such as vulnerable people with low socioeconomic status (SES), are at an increased risk of physical inactivity and respond in diverse ways to specific parts of a PA intervention [22]. For example, previous research highlights the importance of considering significant barriers such as fear of injury and promoting awareness of the benefits of PA in designing PA interventions for older adults [23]. Comparable, it is known that for people with and without PA impairment [9,22], the determinants of PA are different, something that is also likely for people with different levels of education [24]. Understanding the concrete differences between subpopulations with regard to PA behavior change mechanisms would allow tailoring to be optimized in order to improve intervention effects and the sustainability of these effects. Results of a recent study by Tummers et al [25] already give rise to tailored interventions for gender-specific subpopulations. This emphasized the importance of conducting further research into possible other factors defining subpopulations that are relevant for tailoring.

Previous research [25] has revealed knowledge about concrete roles of key determinants in PA behavior change, which is relevant information for designing and tailoring future interventions. In this study, an overview of complex interactions between a PA intervention, determinants, and PA outcome levels was provided for a general population of older adults (50 years and older) and gender-specific subpopulations [25]. The importance and roles of determinants in short-term (6 mo after the baseline) PA behavior change and maintenance of intervention effects in the long term (12 mo after the baseline) have been shown for these groups. In general, habit and intention appeared to be directly related to PA outcome levels. Social concepts, self-efficacy, attitude, intrinsic motivation, commitment, and planning all play a relevant indirect role, affecting PA via other determinants. The models demonstrate that most determinants influence short-term PA, and that these effects permeate into the longer term via previous PA or the direct determinants attitude pros (ie, perceived advantages of PA) and intention. Regarding the differences between males and females, the role of self-efficacy differs across both subpopulations, and attitude is shown to be more relevant for males, while planning turned out to be more important for females. These insights into the working mechanisms of PA interventions have been gained by analyzing an integrated dataset of multiple intervention studies applying a machine learning approach based on Bayesian networks [22,26-30]. Since this methodology provided valuable insights, especially with regard to subpopulation-specific differences, it is important to study the relevance and role of determinants of PA of other subpopulations besides gender, similarly. Potentially relevant moderating factors are age, education level, and PA impairment, as explained below.

Age has been identified in several studies as a moderator of PA behavior change [22,24,31]. In addition, the type of PA behavior has been shown to depend on age [32]; domestic activities make up a greater proportion of total moderate- to vigorous-intensity PA among older adults in contrast to exercise, fitness, and team sport activities. Therefore, differences in age groups are studied. Further, vulnerable groups are often related to low health literacy, making it relevant to see to what extent people's degree of education or their SES affects the PA behavior change mechanisms [33]. Despite mixed findings and varying measurement approaches complicating interpretations and explanations, a review study showed that there is an association between SES and PA [34], where a higher SES is generally associated with greater PA in adolescents. In addition, educational level has been shown to be a moderating factor of intervention effects on PA, examined by analyses for subgroups, where different levels of short-term (household-activity related) PA increases were found for education-based groups of intervention participants [22,28]. Finally, it is important to identify differences in PA behavior change between people with and without PA impairment, as determined by the presence of functional impairments that hinder PA [27]. People experiencing PA restrictions are shown to be less physically active than those without physical impairments [27]. Furthermore, experiencing limitations in physical activities makes it much more difficult for individuals to increase their PA; previous research has shown that personal barriers such as an impairment should be considered when advising on sports and that the presence of physical impairments is a moderator of intervention effects on PA [22,35]. Therefore, it is relevant to look into the specific mechanism for people with any restriction in physical functioning compared with those without PA impairment.

Previous literature has thus shown differences between subpopulations defined by the factors age, education, and PA

XSL•FO
RenderX

impairment; however, little insight remains in how these factors relate to the roles of determinants. Because of that, it is expected that it is relevant to further examine these subpopulations. Therefore, in this paper, we explore differences in PA intervention working mechanisms of the subpopulations based on age, level of education, and being physically limited. This study uses a Bayesian network analysis for each of these subgroups in order to identify differences and similarities in the relevance of determinants and their roles in the short- and long-term per moderating factor. Since the factors studied are shown to be relevant moderators, it is expected to find differences that can provide useful information for optimizing tailoring interventions based on these factors.

Methods

Case Description

This study analyzes a compound dataset of 5 different studies on the effectiveness of PA interventions, previously developed and evaluated, that focus on changing PA behavior through computer-tailored personal advice, selected based on the similarity in the design of the intervention content [22,25,27-30,36]. For completeness, information about the effectiveness and main characteristics of the included studies with respect to intervention design is presented, leading to differences in emphasis of content (Table 1). To get an overall overview, beyond personalized intervention content, the data from multiple studies have been integrated based on study-overarching concept definitions. This study investigates general differences in intervention effects and mechanisms across the defined subpopulations. Thereto, models are estimated for subpopulation-specific subsets of the dataset, defined by a moderating factor.

Table . Characteristics and effectiveness of included intervention studies [25].

Study index (reference)	Intervention target group	Intervention delivery mode	Effectiveness at 6 and 12 months after baseline
1 [18]	Adults aged 50 years and older	Written	Yes. Depends on intervention content (more specifically, only the intervention including environmental information is effective) and differs between age-, BMI-, and intention-specific subgroups.
2 [8]	Adults aged 50 years and older	Written and digital per individual	Yes. Depends on delivery mode and environmental conditions (no intervention effect found in web-based environmental conditions). Also, for some conditions, the moderation effects of age, gender, and intention have been found.
3 [16]	Single, adults aged 65 years and older	Written and digital	No. Effects at 3 months, but these have evaporated at 6 months. No measurement at 12 months.
4 [9]	Adults aged 65 years and older having chronic or physical limitations for PA[a]	Written and digital	Yes, although limited. Depends on the degree of impairment, BMI, age, and educational level.
5 [17]	Former/current patients with cancer aged 18 years and older	Written and digital	Yes. Depends on cancer type, educational level, gender, and age.

[a]PA: physical activity.

In this study, relevant subpopulations based on the factors age, level of education, and PA impairment were defined. A fairly recent study shows that many older adults (older than 65 years) have no intentions to be physically active [37]. Differences in intention and also in lifestyle (eg, working vs retired) led to the decision to distinguish between the subpopulations of participants younger than 65 years and participants being 65 years or older. In order to gain insight into mechanisms for vulnerable groups with low SES, this study focuses on the behavior change of less educated individuals compared with average or higher-educated people. Thereto, the classification for education level defined by the Dutch governmental institution Statistics Netherlands [38] is adhered to in this study. Further, participants who reported having PA impairments are compared with participants who were not physically limited. This study adopts a broad classification of individuals as either being limited in their PA or not, in order to account for variations in underlying chronic conditions and in the degree of physical limitations across individuals and studies. This approach aligns with the focus of the interventions on promoting PA rather than on specific medical conditions. The heterogeneity within the group with PA impairments should be acknowledged, as individuals within this group may have different perceptions towards PA and encompass various underlying conditions. Despite this, they share the common challenge of facing additional physical barriers to perform PA compared with the

XSL•FO
RenderX

other, healthier group, making tailored intervention strategies particularly relevant.

Study Data

As mentioned, an integrated dataset derived from data of the 5 e-health intervention studies is the subject of this study [22,27-30,36]. Effects in such intervention studies are measured by the collection of data from intervention participants as well as participants in a (waiting-list) control group, who have not received any intervention content during the experiment (Table 1 describes the effects). The control group of each study is comparable to the experimental group of the corresponding study, and participants of the control group received intervention content after the experiment phase and final measurements. All participants received questionnaires at different moments in time, namely at the baseline (T0, just before receiving intervention content), and at 3 (T1), 6 (T2, short-term effects), and, in most studies, 12 (T3, long-term effects) months after the baseline [8,9,16-18]. Due to differences in measured variables and measurement timing across studies, missing values emerged during dataset integration. In this paper, analyses are performed on subpopulations, that is, subsets of the integrated dataset including the most important variables that were also analyzed in previous, related research [25]. For each subpopulation for which a Bayesian network is estimated, a relevant subset of records is taken from the integrated dataset, depending on participants' values for the factor that is analyzed. For validation, the numbers of control and intervention group participants for each subpopulation that is analyzed are presented (Table 2).

Table . Number of participants in the control and intervention conditions for each investigated subpopulation group. Due to missing values in the demographic factor variable, the total number of participants (in the control or intervention group) of the 2 subgroups for the different evaluated factors may not match.

Subpopulation	Number of participants	
	Control group	Intervention group
Age (years)		
<65	643	2019
≥65	922	2126
Education		
Low	682	1597
Medium to high	734	2153
Physical activity impairment		
Without	666	2186
With	790	1777

Measurements

The integrated dataset consists of PA outcomes, the main socio-cognitive determinants of PA, demographics, and the intervention variable (Table 3).

The outcome measure of total minutes of moderate- to vigorous-intensity PA per week is central to the analyses in this study. This outcome measure of intervention effects on PA behavior is derived from the raw data of the self-administered Dutch Short Questionnaire to Assess Health-Enhancing PA (SQUASH) included in the questionnaires that participants received [39].

The core determinants are selected based on the studies included in the analysis, requiring that each determinant be measured in at least 2 studies. The underlying intervention studies draw on theory- and evidence-based models, which demonstrated the relevance of these determinants to PA behavior. The selected determinants are self-efficacy, attitude (pros and cons), intrinsic motivation, intention, commitment, strategic planning, action planning, coping planning, habit, social modeling, and social support. The premotivational determinants of self-efficacy refer to an individual's expectation about their own capacity to engage in PA, and attitude reflects an individual's positive and negative evaluation of PA. With respect to the motivational determinants, intrinsic motivation refers to the extent to which an individual engages in PA for personal enjoyment or inherent satisfaction, that is, without external influence, and intention to the extent to which an individual plans or aims to engage in PA. The postmotivational determinant commitment concerns the extent to which an individual is dedicated to maintaining PA over time. With respect to the different (postmotivational) planning determinants, strategic planning involves formulating an explicit approach engaging in PA, action planning refers to specifying when and how to perform PA, and coping planning entails adjusting plans to remain active despite unexpected circumstances. The final postmotivational determinant habit refers to the extent to which PA is performed automatically and consistently as part of an individual's routine. Finally, determinants related to an individual's social environment are included, namely social modeling referring to the influence of observing the behavior of others , where individuals may adopt PA patterns based on examples in their social environment, and social support reflecting the direct encouragement or assistance an individual receives from others, such as family or friends, to engage in PA. Figure 1 shows information about determinant roles from established behavior change models [40-42]. The definitions described above are also derived from these models. As depicted, the determinants interact at different stages of PA behavior change, where premotivational determinants influence

XSL•FO
RenderX

an individual's awareness and readiness to engage in PA, motivational determinants drive the decision to initiate PA, and postmotivational determinants support the maintenance of PA over time by facilitating consistency and overcoming barriers [43]. The determinant concepts mentioned were assessed by several items, and corresponding concept variables are aggregated from the raw item data. Note that the operationalization of concepts differs to some extent across the included studies. All measures were self-reported. The concept scales are computed as the mean of the corresponding unipolar item scales, allowing a maximum of 25% missing items per concept. Except for intention, which was measured on a 10-point scale, items were measured on a 5-point scale, and the number of items per concept varies across the studies and across concepts.

Table . Timeslots of measurement of determinants and outcomes (including indication of studies) [25]. For the subpopulation younger than 65 years, no data are available from studies 3 and 4, and for the subpopulation without impairments, limited data are available from study 4.

Variable/timeslot	Baseline (T0)	3 months (T1)	6 months (T2)	12 months (T3)
Self-efficacy	X (1,2,3,4,5)[a]	X (1,2,3,5)	X (5)	X[b] (4)
Attitude (pros and cons)[c]	X (1,2,3,4,5)	X (1,2,3,5)	X (5)	X (4)
Intrinsic motivation	X (1,2,5)	X (1,2)	X (5)	—[d]
Intention	X (1,2,3,4,5)	X (1,2,3)	X (1,2,4,5)	X (1,2,4,5)
Commitment	X (1,2)	X (1,2)	X (1,2)	—
Strategic planning	X (1,2,3,4,5)	X (1,2)	X (1,2,5)	X (1,4)
Action planning	X (1,2,5)	X (1,2)	X (1,2,5)	—
Coping planning	X (1,2,5)	X (1,2)	X (1,2,5)	—
Habit	X (1,2,4,5)	—	X (1,2,4,5)	X (1,2,4,5)
Social modeling	X (1,2,3,4,5)	X (1,2,3)	X (1,3)	X (4)
Social support	X (1,2,3,4,5)	X (1,2,3,5)	X (3,5)	X (4)
SQUASH[e] outcome (MV-PA)[f]	X (1,2,3,4,5)	X (1,2,3,5)	X (1,2,3,4,5)	X (1,2,4,5)

[a]"X" indicates that a variable was measured at a specific time point. The numbers in parentheses refer to the studies in which it was measured at that time point.

[b]Raw data measurements are very limited.

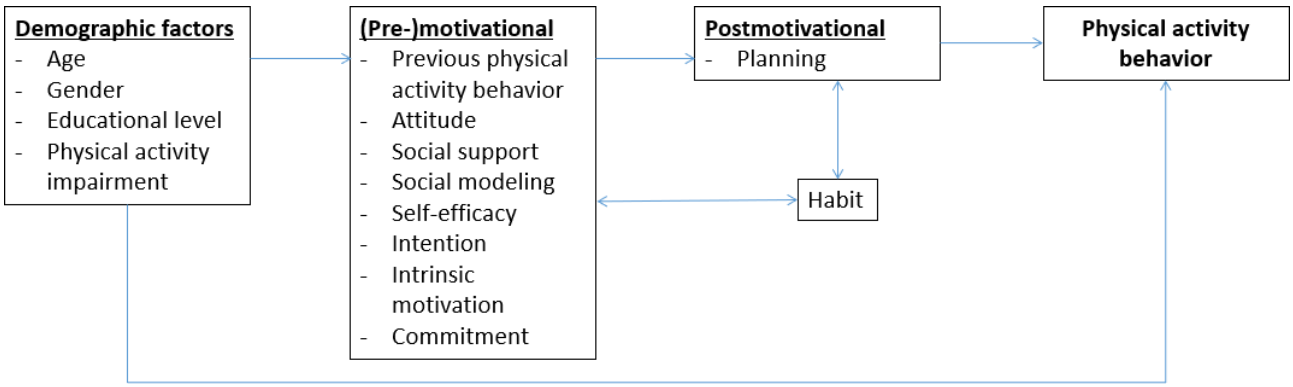
[c]Attitude pros and attitude cons have been measured and are included in analyses as separate concepts.

[d]The concept was not measured at that specific time point.

[e]SQUASH: Short Questionnaire to Assess Health-Enhancing Physical Activity.

[f]MVPA: moderate- to vigorous-intensity physical activity.

Figure 1. Theoretical intervention framework for the role of determinants in influencing physical activity [40-42].



Each of the sociodemographic factors, age, gender, educational level, and PA impairment, is either included in a particular model, or the focus of the analysis if the investigated subpopulations are defined based on the demographic. In analyses where the subpopulation under consideration is not defined based on age, this demographic is uncategorized and measured in years. Gender and PA impairment are included as Boolean variables. Educational level is categorized into low, medium, and high, according to the Dutch education system [38]. In analyses in which this demographic defines the subpopulation under consideration, the medium and high-educated subgroups are merged into a single category.

XSL•FO
RenderX

A Boolean intervention variable is included in analyses in this study, indicating whether a participant was part of the control or the intervention group. It should be noted that the content that intervention participants received is unique for each individual because of computer-tailored advice, as mentioned in the background section. However, in tailoring, the intervention content is altered minimally, based on the answers of respondents to a large number of tailoring questions. Therefore, the personalized intervention content is beyond the scope of this paper but can be found in previous papers [8,9,16-18].

Analysis

In this study, a Bayesian network approach is applied to evaluate differences in PA behavior change mechanisms between subpopulations defined by the demographic factors age, educational level, and PA impairment, respectively. The experiments were conducted in R (version 4.0.3; The R Foundation for Statistical Computing), and the package bnlearn (version 4.6.1) was used for Bayesian network learning [44,45]. Additionally, the package igraph (version 1.2.6) was used to extract desired paths from the Bayesian network models, mpmi (version 0.43.1) for the estimation of mutual information, and Rgraphviz (version 2.34.0) to visualize the Bayesian networks [46-48]. Different models are learned from a subset of the data for which one of the moderating factors has certain values of interest, and this (constant) factor has been left out of the model. By taking a subset of the data, it is possible that for some variables, there is no data at all and has thus been excluded from the analyses of a specific subpopulation. Each model considers a subpopulation defined by one of the demographic factors mentioned and corrected for the other 2 demographic factors and gender.

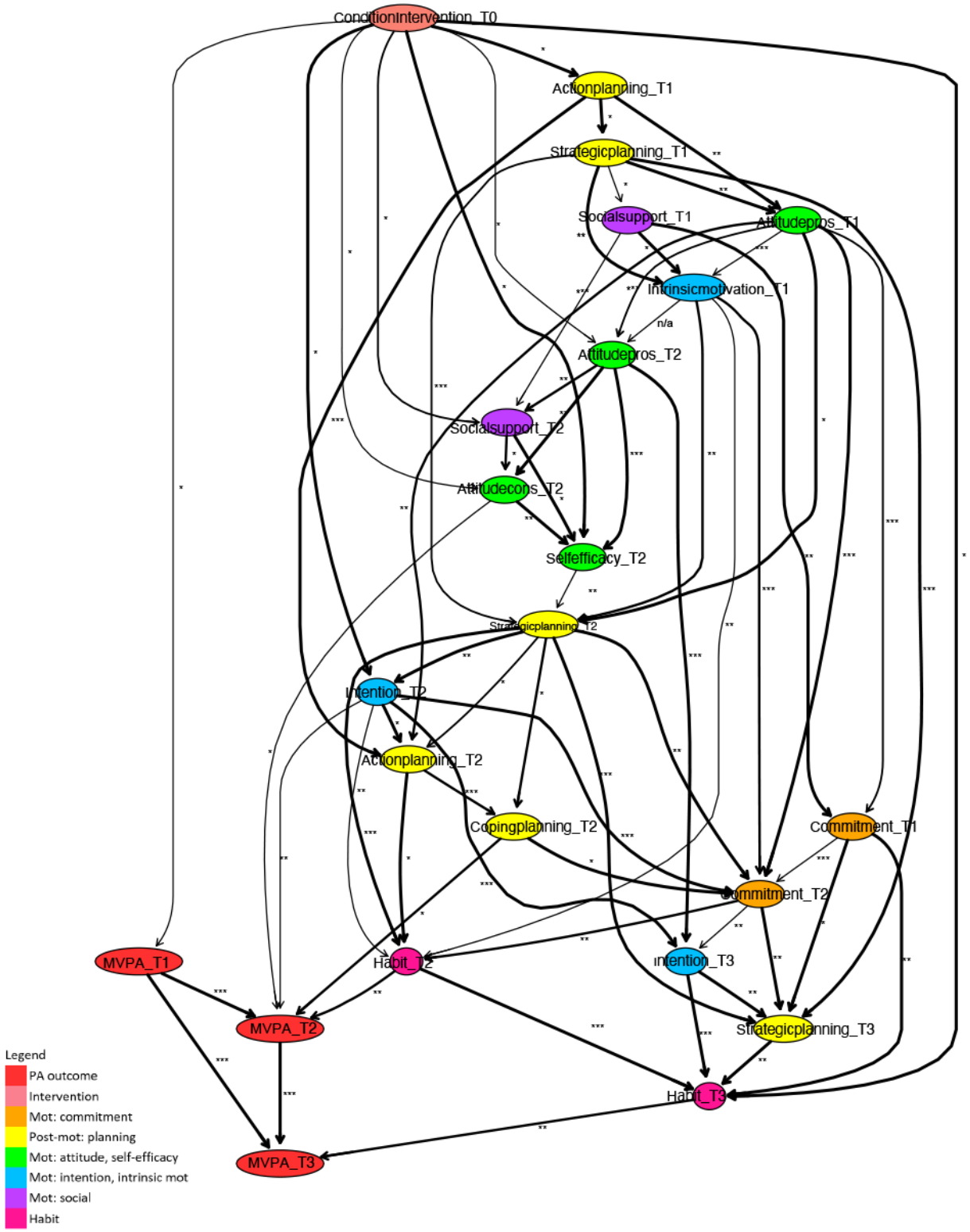
Bayesian networks are probabilistic models that represent relations between variables based on conditional independencies and that can be represented as directed acyclic graphs [49]. The Bayesian network models that are learned are temporal and hybrid, assuming a multinomial and Gaussian distribution for the discrete and continuous random variables, respectively. The networks are restricted to ensure that arcs are consistent with the time dimension of measurements.

To estimate a Bayesian network that includes a stable set of arcs, the same approach as in previous, related articles is used [25], applying Efron's Bootstrap [50]. To explain this approach briefly, we chose the number of bootstrap samples ranging from 100 to 150 based on the structural Hamming distance stability measurement, for a chosen confidence threshold. To learn a Bayesian network model for each bootstrap sample, an algorithm is applied that combines structure learning with the estimation of missing values, called the structural expectation maximization algorithm [51]. The model structure is learned by searching through candidate graphs and selecting the structure that best fits the data according to the Bayesian Information model selection Criterion [52,53], under the time dimension-related restrictions. An averaged model is created based on the bootstrap sample models, in which arcs are selected with at least 60% confidence. It should be noted that, in case both directions occur equally often in the bootstrap sample models, relations in the averaged model are considered undirected.

To draw conclusions from the derived averaged model regarding PA intervention working mechanisms, the relevant paths between the intervention condition variable and the short- and long-term PA outcome variables are deduced. For interpretation purposes, the model is visualized, where the strength of each relation is indicated by asterisks and the stability in the arcs' thickness (Figures 2-7). The asterisks are derived from the jack-knife bias-corrected mutual information estimates, based on complete cases, and relatively allocate arcs to 3 groups by cutting off at 33% and 67% quantiles [54]. Note that the exact cut-off points are given in the results section for each model presented. The different arcs' thickness levels divide arcs according to the percentage of bootstrap sample models in which the specific arc occurs, as described, being at least 0.6, into 4 groups, cutting off at 0.7, 0.8, and 0.9. To simplify the comparison of final models with respect to relevant determinants and their roles in pathways of intervention influences, different colors of nodes indicate determinant categories.

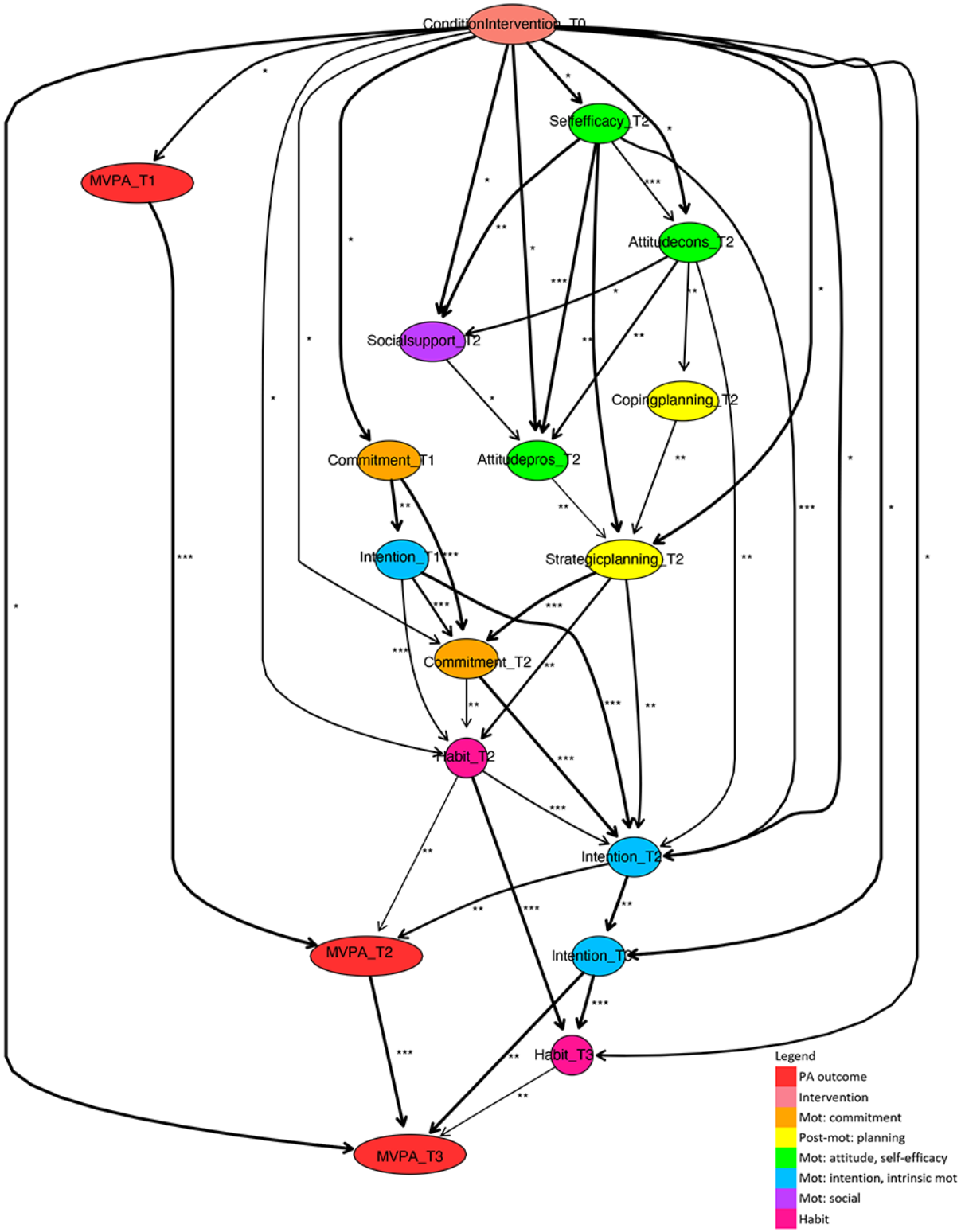
XSL•FO
RenderX

Figure 2. Fragment of a Bayesian network learned for a subpopulation of participants younger than 65 years. Mot: motivation; MVPA: moderate- to vigorous-intensity physical activity; PA: physical activity.



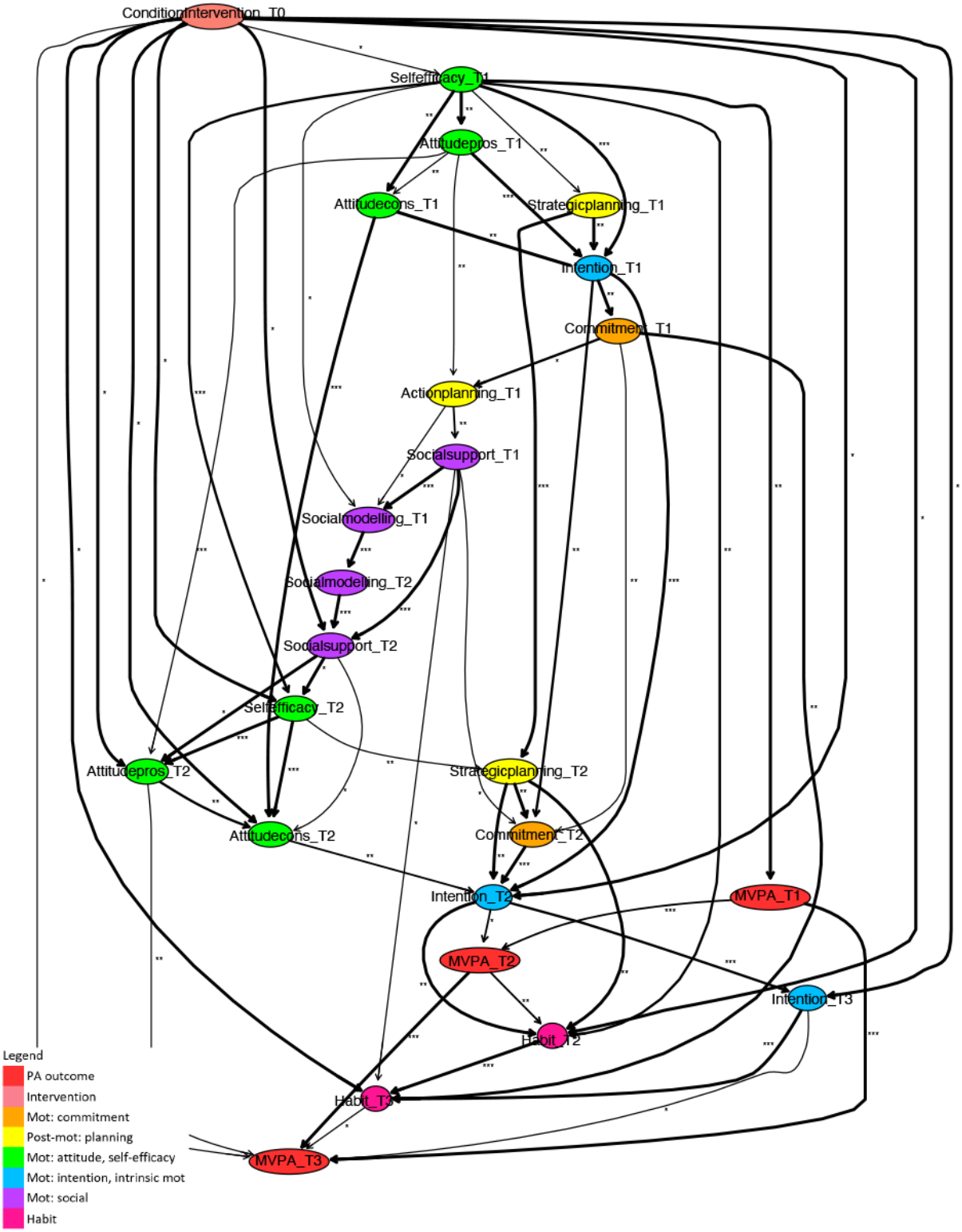
XSL•FO
RenderX

Figure 3. Fragment of a Bayesian network learned for a subpopulation of participants aged 65 years and older. Mot: motivation; MVPA: moderate- to vigorous-intensity physical activity; PA: physical activity.



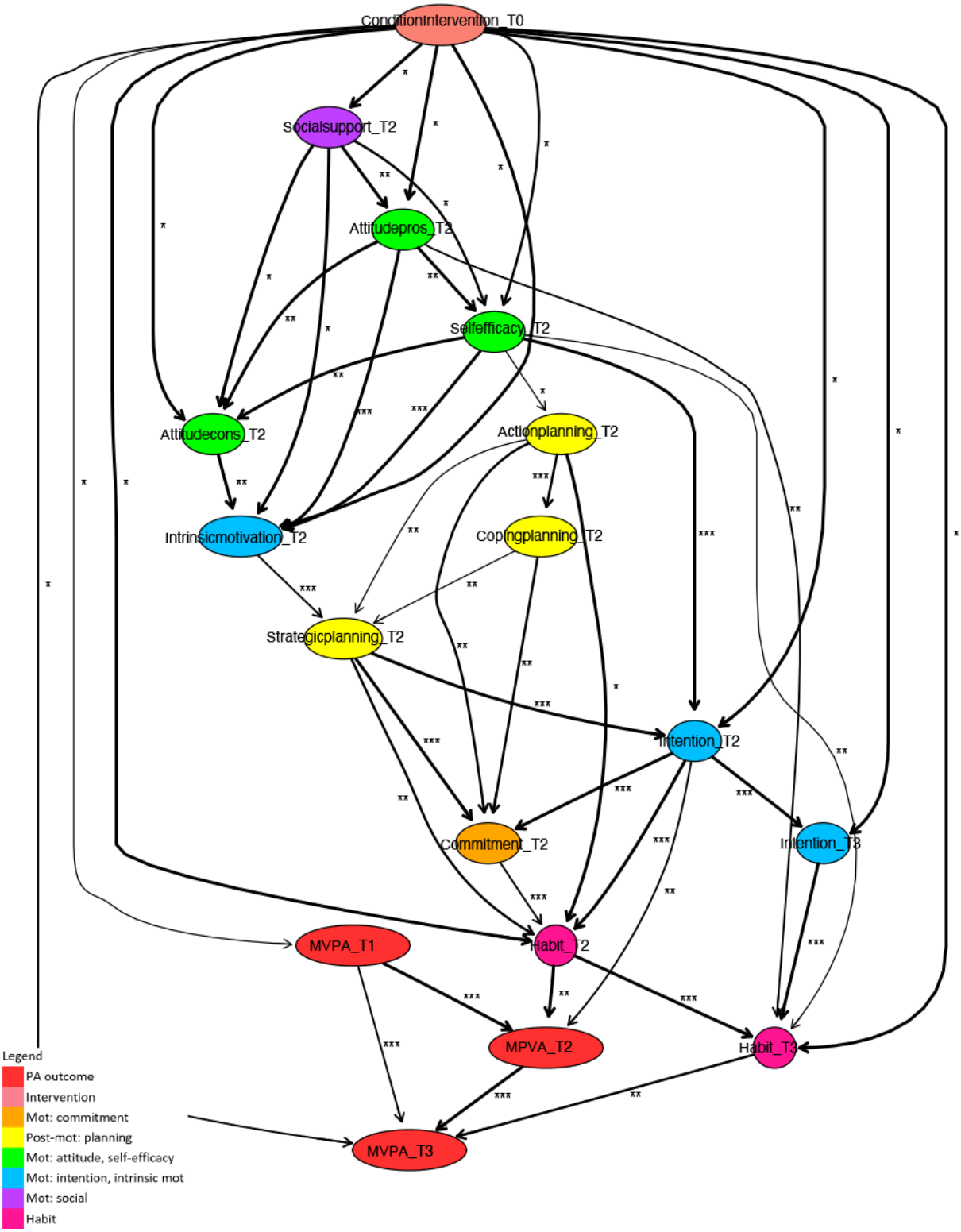
XSL•FO
RenderX

Figure 4. Fragment of a Bayesian network learned for a low-educated subpopulation. Mot: motivation; MVPA: moderate- to vigorous-intensity physical activity; PA: physical activity.



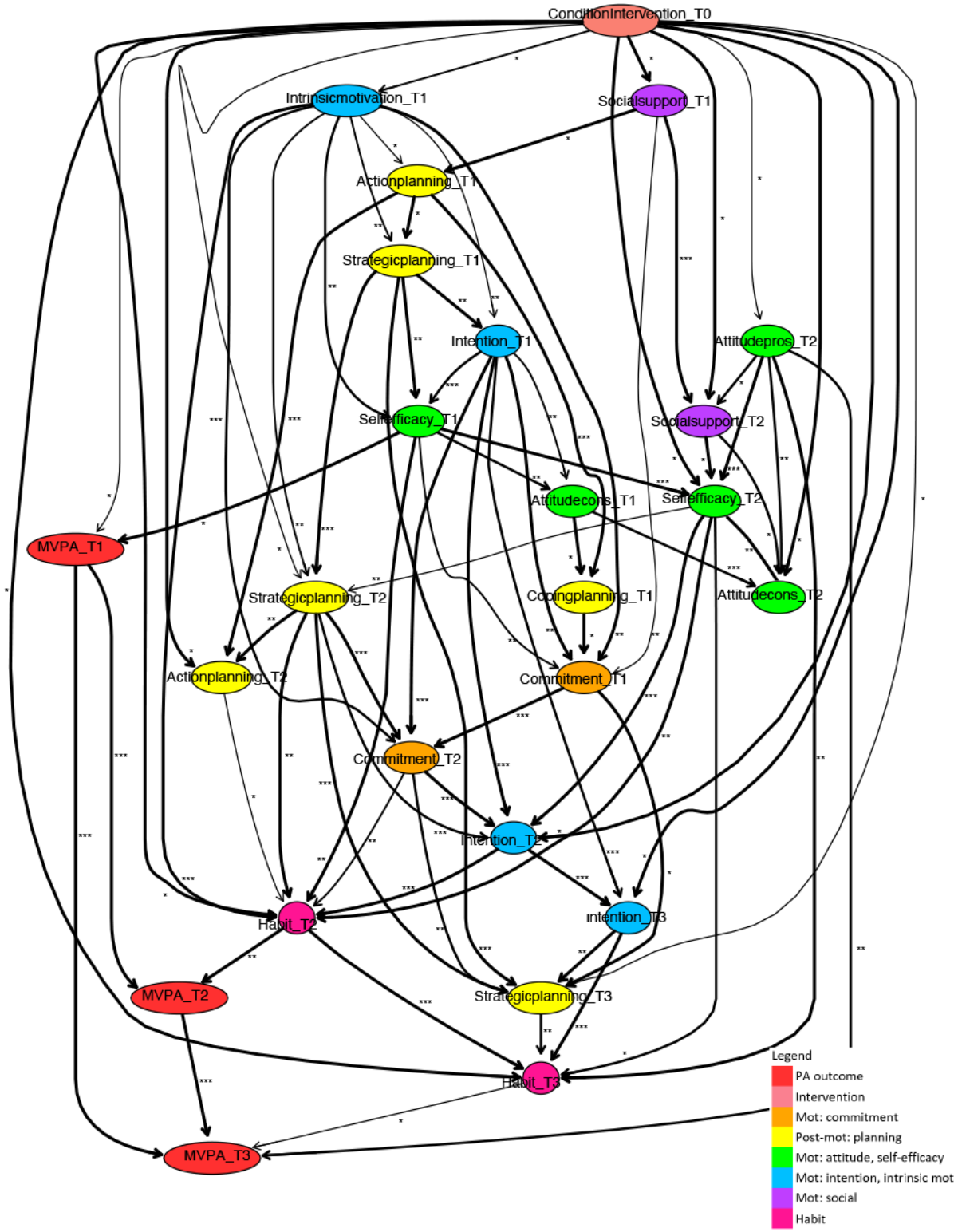
XSL•FO
RenderX

Figure 5. Fragment of a Bayesian network learned for medium- and high-educated subpopulation. Mot: motivation; MVPA: moderate- to vigorous-intensity physical activity; PA: physical activity.



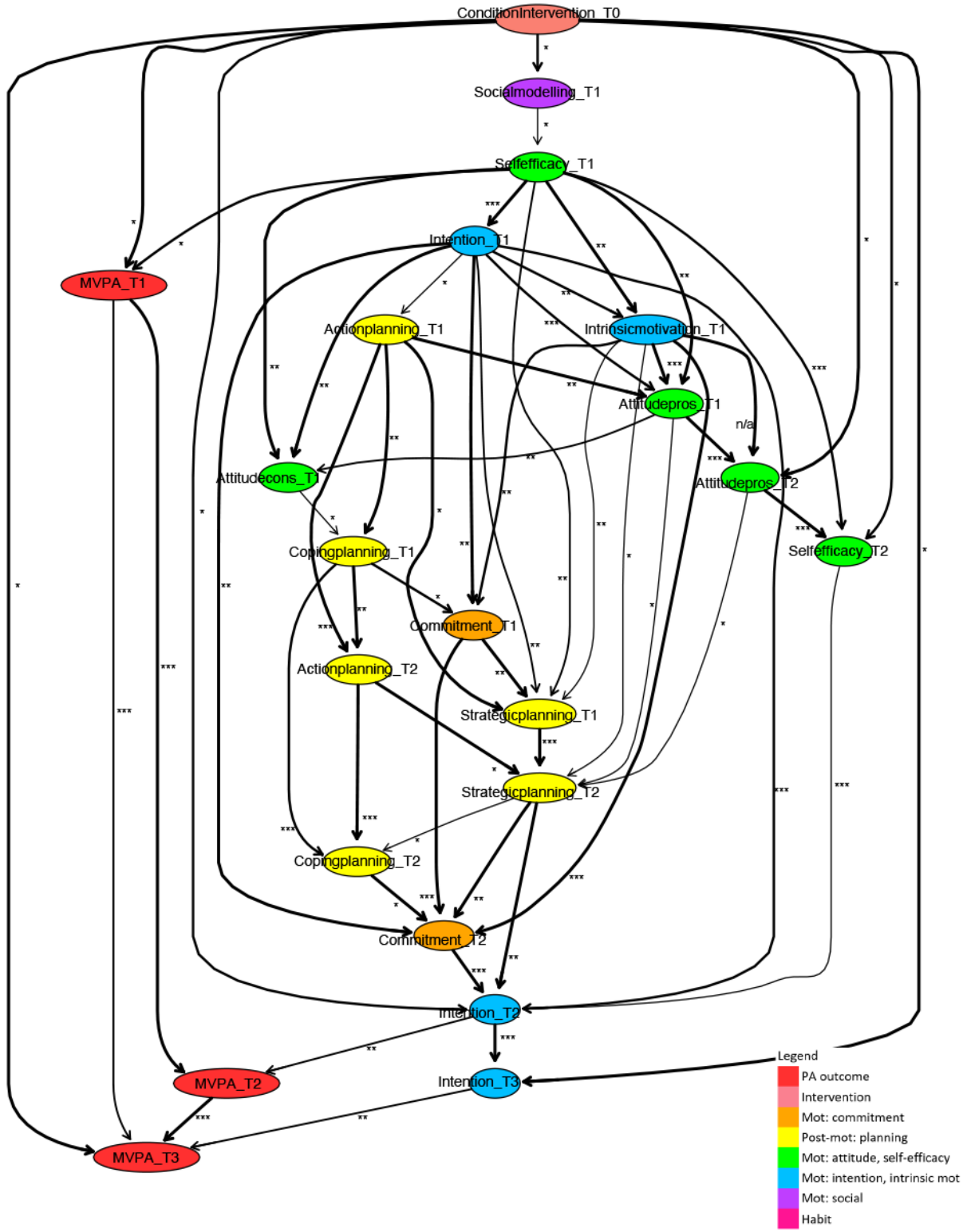
XSL•FO
RenderX

Figure 6. Fragment of a Bayesian network learned for a subpopulation without physical activity impairment. Mot: motivation; MVPA: moderate- to vigorous-intensity physical activity; PA: physical activity.



XSL•FO
RenderX

Figure 7. Fragment of a Bayesian network learned for a subpopulation with physical activity impairment. Mot: motivation; MVPA: moderate- to vigorous-intensity physical activity; PA: physical activity.



Ethical Considerations

The 5 studies in this analysis's dataset [22,25,27-30,36] received ethics approval from the research ethics committee (Commissie Ethische Toetsing Onderzoek) of the Open University of the Netherlands. Institutional review board approval numbers can be found in the original publications. All participants in the original studies provided informed consent. No additional ethics approval was required for the present analysis, as it involves secondary analysis of fully anonymized pre-existing data. According to Dutch national regulations, such research falls outside the scope of the Medical Research Involving Human Subjects Act (WMO) [55]. Furthermore, institutional guidance from the Open University of the Netherlands' research ethics committee (Commissie Ethische Toetsing Onderzoek) confirms that no ethical assessment is required for studies using fully

XSL•FO
RenderX

anonymized secondary data [56]. The authors of the previous studies made the corresponding data available and gave permission for secondary data use.

Results

Overview

This section presents the distilled paths of the averaged Bayesian networks learned from the PA intervention data for subpopulations defined by age, educational level, and PA impairment of participants. First, differences between participants younger than 65 years and 65 years and older are considered. Second, groups of lower and medium-to-high-educated participants are evaluated. Third, PA intervention working mechanisms of participants with PA impairment compared with those without are examined. Thereby, the number of bootstrap samples to learn a model was determined per subpopulation based on the structural Hamming distance (Multimedia Appendix 1). Due to the complexity of the resulting models, the focus of this study is limited to the main similarities and differences that are most profound in comparing the 2 models under consideration for age, education, and physical impairment subgroups, respectively. Since most direct influences of the intervention on determinants are relatively weak, the main effect pathways were selected in the presented submodels based on their stability. Note that influences induced by these weakly affected determinants are relatively stronger, ultimately effectively leading to behavioral change. For the declaration of findings, overviews of the means and SDs of variables in the subpopulation models are available (Multimedia Appendices 2-4).

Age

Overview

Segments of the Bayesian network were learned, similarly following the procedure described in the analysis section, for data of only participants younger than 65 years (Figure 2) and participants aged 65 years and older (Figure 3). The following variables have been removed from the age-based subpopulation sets, since there were no observations for the subpopulation younger than 65 years: self-efficacy at T3, attitude pros at T3, attitude cons at T3, social modeling at T3, and social support at T3. It should be noted that the total integrated dataset consists predominantly of data for older adults, aged 50 years and older.

Short-Term Age-Related Paths

Considering the paths of short-term PA intervention behavior change, for both subpopulations, habit and intention at T2 have a direct relation with short-term (T2, after 6 mo) PA outcomes. In addition, the pathways of intervention effects on these determinants are similar. Indirect determinants that play an important role in these paths are attitude cons, attitude pros, self-efficacy, social support, strategic planning, and commitment at T2.

There are also important differences between the 2 subpopulations. First, attitude cons, and coping planning at T2 have a direct relation with short-term PA outcomes for the subpopulation younger than 65 years, while for the subpopulation aged 65 years and older, they have an indirect influence. Second, for the subpopulation younger than 65 years, action planning at T2 is related to short-term PA indirectly, via habit at T2, while this determinant does not appear in the model for the subpopulation aged 65 years and older. Third, because the direct influence of the intervention on action planning at T1 is highlighted (ie, distilled) in the model for the subpopulation younger than 65 years, several other determinants at T1 appear in this model, while they do not appear in the model for the subpopulation aged 65 years and older. For the younger subpopulation, these determinants at T1 ultimately create effects on the pathways that have been discussed to be similar for both subpopulations, where attitude pros at T1 seem important to predict the sequel.

Long-Term Age-Related Paths

For both subpopulations, most of the intervention effects in the long term originate from short-term pathways. Habit and intention at T3 are ultimately influenced by these pathways and play an important role in long-term effects, where habit at T3 is a direct determinant of PA at T3.

The role of intention at T3 differs between subpopulations. For the subpopulation aged 65 years and older, this is a direct determinant of PA at T3, while for the subpopulation younger than 65 years, the influence of this determinant is indirect and causes effects on PA at T3 via habit at T3. Another difference between the 2 subpopulations is that for the subpopulation younger than 65 years, strategic planning T3 occurs in the long-term pathways, while it does not occur for the subpopulation aged 65 years and older.

Education

Short-Term Education-Related Paths

For both subpopulations (Figure 4, where the edge between attitude cons and intention at T1 is undirected, and 5), habit and intention at T2 are interrelated and both directly related to short-term PA outcomes. These direct determinants are influenced by the intervention directly and indirectly. Within the structure of intervention effects on these determinants, in the models for both subpopulations, the determinants social support, attitude (pros and cons), self-efficacy, commitment, and planning at T2 have similar (indirect) roles.

There are some differences between the 2 subpopulations regarding the occurrence of determinants, as well as the arc direction. First, intrinsic motivation at T2 does not occur in the intervention pathways for lower-educated people, but does occur in those for medium- to high-educated people, where it has an indirect role in the PA intervention behavior change mechanism due to its effect on strategic planning. Second, there is a difference between the 2 subpopulations with regard to which specific planning determinants play a role. At T2, in intervention paths for lower-educated people, only strategic planning is relevant, while for medium- to high-educated people, action planning and coping planning are also important. Third, the direction of the arc between habit and PA at T2 is different, which can be due to the fact that some structures that differ in arc direction model the same probability distribution, that is, they can be Markov equivalent. Fourth, it should be noted that

XSL•FO
RenderX

at first glance, the models suggest large differences between the 2 subpopulations, though in reality, differences may be smaller. Namely, several determinants at T1 occur in the model for lower educated, which ultimately have effects on the corresponding structure discussed earlier (which broadly implies that the intervention influences social and attitudinal determinants, which, potentially through planning and commitment, affect the direct determinants of PA at T2, namely intention and habit) and do not appear in the model for medium- to high-educated. Because the relationship between the intervention and self-efficacy T1 is highlighted, a whole chain of determinants at T1 is also shown in the model. Indeed, it could be the case that this single relation of the intervention and self-efficacy at T1 was just not stable enough for the subpopulation of medium- to high-educated, so that a similar chain of relationship probably present in the whole model has not been highlighted in the submodel.

Long-Term Education-Related Paths

For both educational subpopulations, the models show how long-term intervention effects (ie, at T3) arise from short-term effects, especially through previous PA and because of the permeation of effects on habit and intention into the longer term. Another similarity is that habit at T3 is a direct determinant of PA at T3.

An important long-term difference is that attitude pros at T2 and intention at T3 have a direct influence on PA at T3 for lower-educated people, while for medium- to high-educated people, these determinants have only an indirect influence (via habit at T3).

Physical Activity Impairment

Short-Term Paths Related to Physical Activity Impairment

Although there is a difference in short-term direct determinants (ie, at T2) of the 2 subpopulations (Figure 6, where the edge between self-efficacy and attitude cons at T2 is undirected, and 7), the pathways through which intervention influences arise on these key determinants are roughly similar. Within this structure, self-efficacy and attitude influence planning and commitment, which in turn influence habit or intention. In both models, determinants at T1 also occur. Intrinsic motivation and intention at T1 play an important role because of their direct relation with many of the determinants within the discussed structure.

As already briefly mentioned, there is a difference in direct determinants within PA intervention pathways between the 2 subpopulations. Intention at T2 is a direct determinant of short-term PA for participants with PA impairment, while it has an indirect influence for participants without PA impairment. For participants without PA impairment, habit at T2 is a direct determinant of short-term PA, while this determinant does not occur in the model for participants with PA impairment. Another difference is in direct intervention effects on determinants. For participants without PA impairment, most determinants are directly influenced by the intervention, whereas for participants with PA impairment, only some are directly influenced.

Long-Term Paths Related to Physical Activity Impairment

For both subpopulations, long-term intervention effects arise from short-term pathways, but the role that the determinants play in it mainly differs. The direct determinants for one subpopulation have a more indirect, or even no role, for the other subpopulation. For participants with a PA impairment, intention at T3 is the direct determinant of PA at T3, while it has a more indirect role for participants without PA impairment. For participants without PA impairment, the direct determinants of long-term PA are habit at T3 and attitude pros at T2. For participants with PA impairment, habit at T3 does not occur, and attitude pros at T2 only has an indirect role because of its role in the short-term pathways of intervention effects. Beyond this, strategic planning at T3 has an indirect relation via habit at T3 with long-term PA for participants without PA impairment, while it does not occur at T3 for participants with PA impairment.

Discussion

Principal Findings

Overall, the structure of short-term pathways of intervention effects corresponds to the compared subpopulations, and long-term effects arise from those pathways. Habit and intention are important determinants of short-term PA, as for most subpopulations in the intervention pathways, they have a direct relationship with PA at T2. Only for participants without PA impairment does intention have an indirect role, and for participants with PA impairment, habit is not important. In addition, attitude cons and action planning play a more important role in the pathways for the population younger than 65 years than those for 65 years and older. Further, intrinsic motivation and planning variables play a less important role for the low-educated than for the medium- to high-educated. In summary, determinants play an important role in the short-term pathways of intervention effects for one subpopulation, while they have none or a more indirect role for the opposing subpopulation. With regard to direct intervention influences, there is a direct relation between the intervention and many determinants in most subpopulation models considered, which means that the intervention is capable of changing determinants. However, for the subpopulation of people with PA impairment, few determinants are directly influenced by the intervention. For example, intrinsic motivation, planning concepts, and habit are directly influenced by the intervention for the subpopulation without PA impairment, but not for participants with PA impairment. The upcoming section elaborates on the implications associated with insights from this study.

Methodological and Interpretation Issues

To increase statistical power, which is particularly important in case analyses that focus on subdatasets, data from multiple studies are integrated. Partly stemming from the data integration process, analyses in this study had to deal with substantial amounts of missing data. In dealing with this, a method for dealing with missing data that has been evaluated within a similar context in previous research was applied [57]. Through a bootstrap procedure, we have generated robust outcomes and

XSL•FO
RenderX

gained insights into the stability of relations in the models. Directions of arcs in the resulting models were determined based on bootstrap sample models and temporal restrictions. Few undirected edges were observed in the resulting models, which means that there is uncertainty about the exact direction of certain relations. While this complicates the interpretation of results, it reflects the complexity of a network of interactions in which multiple variables are interconnected.

With a focus on the study objectives, paths have been distilled from the models that also meet a stability threshold, which was determined while considering false positives and false negatives of the identified relationships. These paths provide a comprehensive and reliable overview of the mechanisms of subpopulations. In particular, this study focuses on vulnerable groups such as people with low education, who are known to have low health literacy. Previously, there was limited insight into intervention mechanisms for specific target groups. However, network models are complex in nature, posing a challenge in comparing them and extracting key accent differences. One can potentially derive substantial information from the subpopulation models about how to adapt and tailor interventions in order to enhance their effects. An initial step has been taken in interpreting these models in the context of adaptation. Further research on the impact of adjustments on intervention effects should reveal the optimal approach for translating mechanism models into effective modifications.

Note that different determinants at T1 appear in the model for those younger than 65 years and in the model for the low educated, which do not appear in other subpopulation models. Because of the highlighted fragments from models (ie, not all arcs/edges are visualized), one has to be careful not to get an inaccurate idea about the differences between subpopulations. More specifically, the occurrence or nonoccurrence of a whole pathway of determinants in a fragment, for example, may depend on the decision whether the influence of the intervention on just one of those determinants is stable enough to be highlighted. In view of these and other results, it is worth keeping in mind during interpretation that the models shown represent specific paths and the most stable relations, as part of a larger model. Furthermore, it is important to note that the results in this study have solely compared subpopulations defined based on a single demographic factor. To gain a more in-depth understanding of how to approach tailoring optimally, it might be necessary for future research to examine subpopulations defined based on combinations of demographic factors as well.

Comparison of Results With Prior Work and Their Practical Implications

Overview

In this section, it is explored how the research outcomes from this study relate to existing knowledge in the context of PA behavior change. Note that whereas previous research has zoomed in a little specifically on the working mechanisms for different demographic groups, this study focused on subgroup-specific mechanisms. Based on the described findings about these specific PA intervention mechanisms, indications can be provided on how intervention contents can be tailored to induce and enhance long-term effects for the investigated subgroups. Several key points are highlighted based on the main differences. Note that further accent variations could be derived from the models presented in this study.

Comparison to the General Population

In earlier, related research on PA behavior change induced by interventions for a general population of (mainly) older intervention participants, a similar behavior change structure was found for short-term PA, with the same direct determinants (ie, habit and intention) and long-term effects that follow from this [25]. Beyond these similarities between subpopulations, which are also consistent with findings in a general population, in this paper, important differences are presented for subpopulations defined by age, educational level, and PA impairment. These differences imply that some determinants are more relevant in the paths of intervention effects for certain subpopulations or have a slightly different role in these paths. As with the difference between males and females previously studied [25], the insights into subpopulation-specific differences provide valuable information about how to tailor interventions.

Subpopulations Defined by Age

Previous research emphasizes differentiated approaches in intervention development, highlighting age as a factor that should be considered in tailoring interventions for specific target populations [58,59]. This study has revealed some nuanced differences regarding determinants that play a slightly more prominent role for certain age groups, such as attitude cons and action planning. Furthermore, prior research has indicated that specific groups exhibit distinct motives for PA, such as health aspects for older participants and social aspects for younger participants [58]. The scientific knowledge about differences in motives is valuable in the context of PA determinants. Motives underlie many of the determinants. There are, for example, particular motives that contribute to the creation of PA habit for a specific subgroup. Findings from this study suggest that determinants such as attitude cons and action planning play a more important role in pathways of intervention effects for people younger than 65 years compared with those for people aged 65 years and older. These findings can be explained by the greater flexibility of younger adults to translate intentions into behavior, whereas older adults, who may already have higher motivation, are less likely to rely on motivational or cognitive factors and more likely to rely on habitual behavior [60,61]. For older adults, information on creating and maintaining habits may be particularly beneficial. To tailor interventions age-based and promote PA behavior change in the younger population, it might be more effective to emphasize these determinants. Note that insights about underlying motives are useful in determining how to affect the designated determinants.

Subpopulations Defined by Education

Previous research recommends developing tailored strategies to promote PA among socioeconomically disadvantaged groups, acknowledging the challenges these groups face in sustaining PA [62]. This study provides suggestions on how this can be achieved. For instance, different roles are found for the planning

XSL•FO
RenderX

concepts, indicating that intervention elements that aim to stimulate action and coping planning approaches might be less effective for low-educated people, while improving strategic planning techniques may be as beneficial for low as for medium- to high-educated individuals. These findings can be explained by differences in cognitive processing and self-regulation skills across educational levels [63,64]. Lower-educated individuals may be less accustomed to structured planning techniques, while strategic planning approaches focusing on habit formation appear to be beneficial regardless of education level. In addition, concerning the maintenance of effects in the long term, intention and attitude pros play a more crucial role for low-educated individuals compared with their medium- to high-educated counterparts. Lower-educated individuals may rely more on intentions and attitudes for behavior change due to limited access to resources and less developed self-regulation skills, making their personal motivation more central in determining their behavior. To tailor interventions based on education, focusing on PA of lower-educated people, strategies aiming to enhance the determinants of strategic planning, intention, and attitude pros may be more impactful to enhance PA. Just as with age groups, previous research has provided insights into specific motives as well as barriers related to PA for various social groups [65]. In the implementation of strategies aimed at influencing key determinants identified in this research for the low SES group, one might consider their particular motives and barriers. For example, if financial position is an important barrier, strategies should address affordable alternatives.

Subpopulations Defined by Physical Activity Impairment

Also, for the factor of experiencing restrictions in physical functioning, some research has been conducted on the motives and barriers of relevant subgroups [66]. Previous research has also provided limited insights into determinants of PA for the subgroup of individuals with PA impairment [67], namely that intrinsic motivation and self-efficacy are important. This research confirms this, but also indicates that these determinants play similar roles for individuals without PA impairment. Internal drivers thus facilitate overcoming obstacles to PA. Further, it has been found that habit does not play a role in intervention paths for the group of individuals experiencing PA impairments, probably because PA is more difficult due to experienced impairments and might therefore not be as ingrained in their daily routines, making it less likely to develop into an automatic behavior. More importantly, results in this study show that the intrinsic motivation of participants with PA limitations is not directly influenced by the intervention and that, in general, fewer determinants are directly influenced for the subpopulation with PA impairments. This may indicate that the differences between subgroups based on the moderating factor of the presence of a PA impairment are the greatest, and that advice may not fit the situation of participants with reported physical limitations. In relation to this, a more detailed examination of the degree of PA impairment would be valuable for future research. To tailor interventions in order to enhance the effectiveness of interventions for the subpopulation with PA impairment, it may be beneficial to implement strategies to affect apparently less responsive factors, such as intrinsic motivation and habit, that address their specific needs and challenges related to PA. The effect for this group may be strengthened by, in the intervention design, taking into account intrinsic barriers and key motives such as physical improvement and rehabilitation [66].

Conclusions

This study has focused on understanding the working mechanisms of PA interventions for specific subgroups. It has been shown that PA interventions induce very complex change processes, with the interventions influencing PA behavior through different pathways, both direct and indirect. For the different subgroups, based on age, education, and PA impairment, similarities but also relevant differences in the pathways of behavioral change are found. While further research is certainly important, current insights can be used to make interventions more effective by suiting them even better to the important characteristics of the target population, in particular, vulnerable groups. This study contributes to improved health and well-being within these subpopulations and to reducing health inequalities.

Acknowledgments

This work is part of the research program Active4Life with project number 546003005, which is financed by ZonMw.

Data Availability

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

Authors' Contributions

The research was initiated by LL, CAWB, AH, and RB, and supervised by AH, LL, and CAWB. SCMWT processed the data and performed the analyses. The first draft of the manuscript was written by SCMWT. Versions of the manuscript were reviewed by LL, CAWB, AH, and RB, and corresponding revisions were mainly made by SCMWT. All authors read, modified, and approved the final manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Stabilization of averaged Bayesian network subpopulation models during the bootstrap procedure.
[PNG File, 288 KB - ojphi_v17i1e57977_app1.png]

Multimedia Appendix 2
Overview of the mean (SD) of variables appeared in age-based subpopulation models. Because the response rate differs across groups and analyses are based on available data, the n value may differ for the different analyses
[DOCX File, 15 KB - ojphi_v17i1e57977_app2.docx]

Multimedia Appendix 3
Overview of the mean (SD) of variables appeared in education-based subpopulation models. Because the response rate differs across groups and analyses are based on available data, the n value may differ for the different analyses.
[DOCX File, 15 KB - ojphi_v17i1e57977_app3.docx]

Multimedia Appendix 4
Overview of the mean (SD) of variables appeared in physical activity impairment-based subpopulation models. Because the response rate differs across groups and analyses are based on available data, the n value may differ for the different analyses.
[DOCX File, 19 KB - ojphi_v17i1e57977_app4.docx]

References

1. Global Recommendations on Physical Activity for Health: World Health Organization; 2010.
2. Bullard T, Ji M, An R, Trinh L, Mackenzie M, Mullen SP. A systematic review and meta-analysis of adherence to physical activity interventions among three chronic conditions: cancer, cardiovascular disease, and diabetes. BMC Public Health 2019 May 24;19(1):636. [doi: 10.1186/s12889-019-6877-z] [Medline: 31126260]
3. Marques A, Sarmento H, Martins J, Saboga Nunes L. Prevalence of physical activity in European adults — compliance with the World Health Organization's physical activity guidelines. Prev Med 2015 Dec;81:333-338. [doi: 10.1016/j.ypmed.2015.09.018]
4. Taylor RS, Brown A, Ebrahim S, et al. Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. Am J Med 2004 May 15;116(10):682-692. [doi: 10.1016/j.amjmed.2004.01.009] [Medline: 15121495]
5. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809. [doi: 10.1503/cmaj.051351] [Medline: 16534088]
6. Weggemans RM, Backx FJG, Borghouts L, et al. The 2017 Dutch physical activity guidelines. Int J Behav Nutr Phys Act 2018 Jun 25;15(1):58. [doi: 10.1186/s12966-018-0661-9] [Medline: 29940977]
7. Zanuso S, Jimenez A, Pugliese G, Corigliano G, Balducci S. Exercise for the management of type 2 diabetes: a review of the evidence. Acta Diabetol 2010 Mar;47(1):15-22. [doi: 10.1007/s00592-009-0126-3] [Medline: 19495557]
8. Peels DA, van Stralen MM, Bolman C, et al. Development of web-based computer-tailored advice to promote physical activity among people older than 50 years. J Med Internet Res 2012 Mar 2;14(2):e39. [doi: 10.2196/jmir.1742] [Medline: 22390878]
9. Volders E, Bolman CAW, de Groot RHM, Lechner L. The effect of Active Plus, a computer-tailored physical activity intervention, on cognitive functioning of elderly people with chronic illness(es) - study protocol for a randomized controlled trial. BMC Public Health 2019 Aug 30;19(1):1197. [doi: 10.1186/s12889-019-7517-3] [Medline: 31470829]
10. Golsteijn R. Physical Activity Promotion in Prostate and Colorectal Cancer Patients and Survivors: Development and Evaluation of the Computer-Tailored OncoActive Intervention: Open Universiteit; 2019.
11. Deci EL, Ryan RM. Self-determination theory: when mind mediates behavior. J Mind Behav 1980;1(1):33-43 [FREE Full text]
12. Ajzen I. From intentions to actions: a theory of planned behavior. In: Action Control: Springer; 1985:11-39. [doi: 10.1007/978-3-642-69746-3_2]
13. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, Schaalma HC, Markham CC. Planning Health Promotion Programs: An Intervention Mapping Approach: Jossey-Bass; 2006.
14. Brug J, Assema P, Lechner L. Gezondheidsvoorlichting En Gedragsverandering Een Planmatige Aanpak [Book in Dutch]: Koninklijke Van Gorcum BV; 2007.
15. Bauman AE, Sallis JF, Dzewaltowski DA, Owen N. Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. Am J Prev Med 2002 Aug;23(2 Suppl):5-14. [doi: 10.1016/s0749-3797(02)00469-5] [Medline: 12133733]
16. Boekhout JM, Peels DA, Berendsen BA, Bolman CA, Lechner L. An eHealth intervention to promote physical activity and social network of single, chronically impaired older adults: adaptation of an existing intervention using intervention mapping. JMIR Res Protoc 2017 Nov 23;6(11):e230. [doi: 10.2196/resprot.8093] [Medline: 29170146]

XSL•FO
RenderX

17. Golsteijn RHJ, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Development of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: OncoActive. BMC Cancer 2017 Jun 26;17(1):446. [doi: 10.1186/s12885-017-3397-z] [Medline: 28651586]
18. van Stralen MM, Kok G, de Vries H, Mudde AN, Bolman C, Lechner L. The active plus protocol: systematic development of two theory- and evidence-based tailored physical activity interventions for the over-fifties. BMC Public Health 2008 Dec 4;8:399. [doi: 10.1186/1471-2458-8-399] [Medline: 19055806]
19. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [doi: 10.1016/s0738-3991(98)00127-x] [Medline: 10223015]
20. Kok G, Gottlieb NH, Peters GJY, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312. [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
21. Kreuter M, Farrell D, Olevitch L, Brennan L, Rimer BK. Tailoring Health Messages: Customizing Communication with Computer Technology: Taylor and Francis; 2000:270. [doi: 10.4324/9781410603319]
22. Volders E, Bolman CAW, de Groot RHM, Verboon P, Lechner L. The effect of active plus, a computer-tailored physical activity intervention, on the physical activity of older adults with chronic illness(es)-a cluster randomized controlled trial. Int J Environ Res Public Health 2020 Apr 10;17(7):2590. [doi: 10.3390/ijerph17072590] [Medline: 32290072]
23. Arazi H, Izadi M, Kabirian H. Interactive effect of socio-eco-demographic characteristics and perceived physical activity barriers on physical activity level among older adults. Eur Rev Aging Phys Act 2022 Mar 29;19(1):8. [doi: 10.1186/s11556-022-00288-y] [Medline: 35350982]
24. van Stralen MM, de Vries H, Bolman C, Mudde AN, Lechner L. Exploring the efficacy and moderators of two computer-tailored physical activity interventions for older adults: a randomized controlled trial. Ann Behav Med 2010 May;39(2):139-150. [doi: 10.1007/s12160-010-9166-8] [Medline: 20182833]
25. Tummers SCMW, Hommersom A, Lechner L, Bemelmans R, Bolman CAW. Determinants of physical activity behaviour change in (online) interventions, and gender-specific differences: a Bayesian network model. Int J Behav Nutr Phys Act 2022 Dec 19;19(1):155. [doi: 10.1186/s12966-022-01381-2] [Medline: 36536443]
26. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference: Morgan Kaufmann Publishers; 1988:552. [doi: 10.1016/B978-0-08-051489-5.50008-4]
27. Boekhout JM, Berendsen BAJ, Peels DA, Bolman CAW, Lechner L. Evaluation of a computer-tailored healthy ageing intervention to promote physical activity among single older adults with a chronic disease. Int J Environ Res Public Health 2018 Feb 15;15(2):346. [doi: 10.3390/ijerph15020346] [Medline: 29462862]
28. Golsteijn RHJ, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Oct 30;15(1):106. [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
29. Peels DA, Bolman C, Golsteijn RHJ, et al. Long-term efficacy of a printed or a web-based tailored physical activity intervention among older adults. Int J Behav Nutr Phys Act 2013 Sep 2;10:104. [doi: 10.1186/1479-5868-10-104] [Medline: 24228862]
30. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. Efficacy of two tailored interventions promoting physical activity in older adults. Am J Prev Med 2009 Nov;37(5):405-417. [doi: 10.1016/j.amepre.2009.07.009] [Medline: 19840695]
31. Golsteijn R. Physical Activity Promotion in Prostate and Colorectal Cancer Patients and Survivors: Development and Evaluation of the Computer-Tailored OncoActive: Open Universiteit; 2019.
32. Bélanger M, Townsend N, Foster C. Age-related differences in physical activity profiles of English adults. Prev Med 2011;52(3-4):247-249. [doi: 10.1016/j.ypmed.2011.02.008] [Medline: 21338622]
33. Stormacq C, Van den Broucke S, Wosinski J. Does health literacy mediate the relationship between socioeconomic status and health disparities? integrative review. Health Promot Int 2019 Oct 1;34(5):e1-e17. [doi: 10.1093/heapro/day062] [Medline: 30107564]
34. Stalsberg R, Pedersen AV. Effects of socioeconomic status on the physical activity in adolescents: a systematic review of the evidence. Scandinavian Med Sci Sports 2010 Jun;20(3):368-383. [doi: 10.1111/j.1600-0838.2009.01047.x]
35. Jaarsma EA, Dijkstra PU, Geertzen JHB, Dekker R. Barriers to and facilitators of sports participation for people with physical disabilities: a systematic review. Scandinavian Med Sci Sports 2014 Dec;24(6):871-881. [doi: 10.1111/sms.12218]
36. Peels DA, Verboon P, van Stralen MM, et al. Motivational factors for initiating and maintaining physical activity among adults aged over fifty targeted by a tailored intervention. Psychol Health 2020 Oct 2;35(10):1184-1206. [doi: 10.1080/08870446.2020.1734202]
37. Alley SJ, Schoeppe S, Rebar AL, Hayman M, Vandelanotte C. Age differences in physical activity intentions and implementation intention preferences. J Behav Med 2018 Jun;41(3):406-415. [doi: 10.1007/s10865-017-9899-y] [Medline: 29116569]
38. Centraal Bureau voor de Statistiek. URL: https://www.cbs.nl/nl-nl/nieuws/2023/48/hoge-tevredenheid-ondanks-lage-materiele-welvaart-in-caribisch-nederland/opleidingsniveau [accessed 2020-08-15]
39. Wendel-Vos GCW, Schuit AJ, Saris WHM, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [doi: 10.1016/s0895-4356(03)00220-8] [Medline: 14680666]

XSL•FO
RenderX

40. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. Psychol Behav Sci Int J 2017;2(2):555-585. [doi: 10.19080/PBSIJ.2017.02.555585]
41. Vohs KD, Baumeister RF. Understanding self-regulation. In: Handbook of Self-Regulation: Guilford Press; 2004:19.
42. Weinstein ND. The precaution adoption process. Health Psychol 1988;7(4):355-386. [doi: 10.1037//0278-6133.7.4.355] [Medline: 3049068]
43. Sutton S. Stage theories of health behaviour. In: Predicting Health Behavior: Open University Press; 2005:223-275.
44. Scutari M. Package "bnlearn": Bayesian network structure learning, parameter learning and inference: R package version 3; 2012.
45. Scutari M. Learning Bayesian networks with the bnlearn R package. J Stat Softw 2010;35:1-22. [doi: 10.18637/jss.v035.i03]
46. Pardy C. Mutual information as an exploratory measure for genomic data with discrete and continuous variables [Dissertation]. : The University of New South Wales; 2013 URL: https://unsworks.unsw.edu.au/entities/publication/5cef42f5-1b16-4b6c-8aed-38e7f0a4ca8f [accessed 2025-08-11]
47. Csardi G, Nepusz T. The igraph software package for complex network research. Complex Syst 2006;1695:1-9 [FREE Full text]
48. Long L, Gentleman R, Falcon S, Hahne F, Sarkar D. Rgraphviz: provides plotting capabilities for r graph objects: R package version 2500; 2024.
49. Murphy KP. Dynamic bayesian networks: representation, inference and learning [Dissertation]. : University of California; 2002.
50. Friedman N, Goldszmidt M, Wyner A. Data analysis with bayesian networks: a bootstrap approach. Presented at: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence; Jul 30 to Aug 1, 1999; Stockholm, Sweden.
51. Friedman N, editor. The Bayesian structural EM algorithm. Presented at: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence; Jul 24-26, 1998; Madison Wisconsin.
52. Koller D, Friedman N. Probabilistic Graphical Models: Principles and Techniques: MIT Press; 2009.
53. Schwarz G. Estimating the dimension of a model. Ann Statist 1978;6(2):461-464. [doi: 10.1214/aos/1176344136]
54. Pardy C, Galbraith S, Wilson SR. Integrative exploration of large high-dimensional datasets. Ann Appl Stat 2018;12(1):178-199. [doi: 10.1214/17-AOAS1055]
55. Your research: is it subject to the WMO or not? CCMO. URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not [accessed 2025-06-20]
56. Secondary data research [Web page in Dutch]. Open University. URL: https://www.ou.nl/documents/40554/4197127/III_01_Onderzoek_secundaire_data_14122016.pdf [accessed 2025-08-11]
57. Tummers S, Hommersom A, Lechner E, Bolman C, Bemelmans R. Gaining insight into determinants of physical activity using bayesian network learning. In: Communications in Computer and Information Science: Springer; 2021, Vol. 1398:172-187. [doi: 10.1007/978-3-030-76640-5_11]
58. Ley C. Participation motives of sport and exercise maintainers: influences of age and gender. Int J Environ Res Public Health 2020 Oct 26;17(21):7830. [doi: 10.3390/ijerph17217830] [Medline: 33114636]
59. Dorner TE, Stronegger WJ, Hoffmann K, Stein KV, Niederkrotenthaler T. Socio-economic determinants of health behaviours across age groups: results of a cross-sectional survey. Wien Klin Wochenschr 2013 May;125(9-10):261-269. [doi: 10.1007/s00508-013-0360-0] [Medline: 23579882]
60. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
61. Gardner B, Lally P, Wardle J. Making health habitual: the psychology of "habit-formation" and general practice. Br J Gen Pract 2012 Dec;62(605):664-666. [doi: 10.3399/bjgp12X659466] [Medline: 23211256]
62. Craike M, Wiesner G, Hilland TA, Bengoechea EG. Interventions to improve physical activity among socioeconomically disadvantaged groups: an umbrella review. Int J Behav Nutr Phys Act 2018 May 15;15(1):43. [doi: 10.1186/s12966-018-0676-2] [Medline: 29764488]
63. Last BS, Lawson GM, Breiner K, Steinberg L, Farah MJ. Childhood socioeconomic status and executive function in childhood and beyond. PLoS One 2018;13(8):e0202964. [doi: 10.1371/journal.pone.0202964] [Medline: 30142188]
64. Duckworth AL, Quinn PD, Tsukayama E. What No Child Left Behind leaves behind: the roles of IQ and self-control in predicting standardized achievement test scores and report card grades. J Educ Psychol 2012 May 1;104(2):439-451. [doi: 10.1037/a0026280] [Medline: 24072936]
65. Pedersen MRL, Hansen AF, Elmose-Østerlund K. Motives and barriers related to physical activity and sport across social backgrounds: implications for health promotion. Int J Environ Res Public Health 2021 May 28;18(11):5810. [doi: 10.3390/ijerph18115810] [Medline: 34071630]
66. Ascondo J, Martín-López A, Iturricastillo A, et al. Analysis of the barriers and motives for practicing physical activity and sport for people with a disability: differences according to gender and type of disability. Int J Environ Res Public Health 2023 Jan 11;20(2):1320. [doi: 10.3390/ijerph20021320] [Medline: 36674077]
67. Brandenbarg P, Krops LA, Seves BL, et al. Psychosocial factors of physical activity among people with disabilities: prospective cohort study. Rehabil Psychol 2023 May;68(2):164-173. [doi: 10.1037/rep0000488] [Medline: 36780269]

XSL•FO
RenderX

Abbreviations

PA: physical activity
SQUASH: Short Questionnaire to Assess Health-Enhancing Physical Activity

Edited by E Mensah; submitted 15.03.24; peer-reviewed by A Samuel, A Woodward, X Zhao; revised version received 06.04.25; accepted 08.05.25; published 03.09.25.

Please cite as:
Tummers SCMW, Hommersom A, Lechner L, Bemelmans R, Bolman CAW
Bayesian Network Analysis of Intervention-Induced Physical Activity Behavior Change: Comparative Modeling Study Across Age, Education, and Activity Impairment Subgroups
Online J Public Health Inform 2025;17:e57977
URL: https://ojphi.jmir.org/2025/1/e57977
doi:10.2196/57977

© Simone Catharina Maria Wilhelmina Tummers, Arjen Hommersom, Lilian Lechner, Roger Bemelmans, Catherine Adriana Wilhelmina Bolman. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 3.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reconsidering Trust and Information Engagement and Unpacking the Role of Emotion in Public Responses During the Early Stage of a Public Health Crisis in China: Web-Based Survey Study

Zhiming Liu[1], MSc; Jiawei Tu[1], MS; Tien-Tsung Lee[1], PhD; Lu Wei[2,3], PhD

[1]Department of Communication, Faculty of Social Sciences, University of Macau, Taipa, Macau, China

[2]Communication University of Zhejiang, Hangzhou, China

[3]College of Media and International Culture, Zhejiang University, Hangzhou, China

Corresponding Author:
Jiawei Tu, MS
Department of Communication
Faculty of Social Sciences
University of Macau
E21B Humanities and Social Sciences Building
Avenida da Universidade
Taipa, Macau
China
Phone: 853 88228387
Email: jiawei.tu@connect.um.edu.mo

Abstract

Background: The COVID-19 pandemic continues to offer valuable insights into crisis management and risk communication, particularly through retrospective analyses that allow a more comprehensive understanding. Emotional responses played a crucial role in shaping how individuals processed information and built trust in different objects in the early stages of the COVID-19 pandemic.

Objective: This study aimed to investigate how negative emotions influence online information engagement and trust in 4 distinct entities: government, scientists, health care providers, and other people (relatives, friends, family, and strangers).

Methods: A nationwide survey was conducted in China from January 31 to February 9, 2020, involving 1568 adult participants. The data collection was particularly valuable due to the limited access to national samples in China during the early stages of the public health crisis. Participants were asked questions related to negative emotions, engagement with online information, and their trust in 4 different entities (government, scientists, other people, and health care providers) during the pandemic. Mediation analyses were performed to test the associations between the examined variables. A 95% bootstrap CI approach was used to estimate the mediation effects.

Results: This study reveals that negative emotions not only had a direct effect on trust but also indirectly fostered trust in the government and scientists through increased information engagement. There was a positive association (B=0.219, SE 0.023; *P*<.001) between negative emotions and information engagement. In addition, individuals experiencing more negative emotions tended to trust more in the government (B=0.191, SE 0.022; *P*<.001) and scientists (B=0.184, SE 0.017; *P*<.001). However, this effect did not extend to trust in health care providers or interpersonal trust.

Conclusions: The research findings reveal that while negative emotions directly and indirectly enhanced trust in the government and scientists through increased information engagement, they did not significantly impact trust in health care providers or interpersonal relationships in the Chinese context. These findings highlight the different pathways through which emotions and information behaviors affect trust during public health crises, offering critical lessons for future public health emergencies and risk communication.

(Online J Public Health Inform 2025;17:e77790) doi:10.2196/77790

KEYWORDS

risk communication; trust; negative emotions; information engagement; public health crisis

XSL•FO
RenderX

Introduction

Background

The COVID-19 pandemic, spanning from its initial outbreak in December 2019 to its gradual containment, has left an indelible mark all over the world in various aspects of human lives [1,2]. A substantial body of research has continued to revisit the impacts, evaluate preventive strategies, and reflect on the effectiveness of risk communication approaches [3,4].

During the onset of the COVID-19 outbreak in China, the populace grappled with heightened negative emotions triggered by the uncertainties and challenges posed by the deadly virus [5,6]. Quarantine measures confined individuals to their homes, and their access to information and interaction with others was primarily online [7,8]. Such isolation not only altered how people obtained and used information but also influenced their perceptions of societal dynamics at large, fostering a climate characterized by such negative emotions as fear and anxiety [6,9].

While facing a highly contagious and deadly virus, people had to make critical decisions on whom and what to trust [3,10,11]. Thus, investigating how to foster trust is inevitable for risk communication and crisis management [1,3,12]. While previous studies have investigated the role of emotions and information use in a pandemic [5,9], this study, drawing upon the risk information seeking and processing (RISP) model [13], delves into the complex interplay between psychological states, information engagement behaviors, and their potential effects on trust toward different and significant entities including the government, scientists, health care providers, and other people to analyze the factors that influence trust in the initial outbreak of a crisis in China.

Literature Review

Trust and Its Consequences

Trust can be defined as the conviction (how) in the reliability of another person (who) regarding a specific issue (what) that arises in the face of uncertainties [14-16]. Trust plays a pivotal role in society, particularly in risk communication and crisis management, as it directly shapes behavioral compliance during times of uncertainty [17]. Public trust serves as a critical factor in facilitating or undermining cooperative intentions toward authorities while also sustaining confidence in institutional structures, governance systems, and response mechanisms [3,12]. Therefore, understanding the mechanisms and conditions under which trust is formed or eroded is essential for improving crisis response, designing more resilient communication strategies, and promoting behavioral compliance for future emergencies.

Research has revealed that trust can lead to cooperative beliefs and behaviors when faced with uncertainties [18,19]. Survey studies uncovered a positive connection between trust and social distancing and self-care behavior [3,10]. In addition an online survey from 11 countries found that people's evaluation of the severity of the COVID-19 pandemic was influenced by their trust in the news media and their perception of social media bias [18]. During the COVID-19 pandemic in the United States, individuals who trusted right-leaning more than left-leaning media engaged in notably fewer preventive actions [19].

However, how trust in different entities develops and functions during the early stages of a public crisis varies across social settings and leads to diverse understandings. China—widely discussed in crisis management yet understudied regarding its social trust for the emerging phase of crisis when fear and ignorance prevail—presents both theoretical and practical significance for this investigation.

Types and Objects of Trust

Factors such as different forms of information resources can influence *types* and *objects* of trust [20-22]. The object of trust can be another person, a group of people, an organization, an institution, a geopolitical region, or the entire world [23]. Two types of trust, *interpersonal* and *impersonal* trust are conceptualized differently [14,24,25]. *Interpersonal* trust which can be defined as trust in other persons, like familiar people like family members is important [26,27]. *Interpersonal* trust can shape people's inclinations to action, particularly during a crisis or a pandemic [28,29]. *Impersonal* trust is also known as institutional [30-32] or systems trust [20,33,34]. It can be characterized as the expectation or confidence that social systems or institutions would act morally and uphold norms [27]. People's attitudes toward health-related policies issued by social institutions (eg, the government) or systems (eg, health care authorities) largely depend on public trust in the entities that enact the policies [35-37].

Types of trust take on a different meaning and inclinations in a Chinese setting. The Chinese tend to have a higher level of trust in institutions and organizations, such as the government or authoritative figures such as scientists [37-40]. Research reveals that a high level of trust in authorities and experts can facilitate the acceptance of preventive measures, reduce panic, and enhance the effectiveness of risk communication strategies [35,36]. Conversely, distrust can lead to misinformation, resistance, and harmful behaviors [41]. However, the Chinese have a comparatively lower level of trust in interpersonal relationships, particularly with strangers, which hinders interpersonal cooperation as well as health-related behavior, particularly social distancing [41,42]. Both scientists (health experts) and health care providers can play a vital role in communicating health-related knowledge and information in everyday life [41,42]. However, when the public is confronted with a health crisis, the dynamics of trust in scientists versus health care providers can vary [43]. For instance, physicians who directly treat the public and offer suggestions are generally more trusted by Americans than scientists in related fields [44,45].

Emotions, Risk Information Seeking, and Trust

The RISP model provides a theoretical framework for understanding how individuals respond to crises and risks through cognitive (systematic) and affective (heuristic) processes [13]. It emphasizes that in the face of uncertainty, people engage in information-seeking and processing behaviors influenced by factors such as emotions, perceived risk, informational

XSL•FO
RenderX

insufficiency, and social norms [46,47]. This study builds upon and extends the RISP model, examining how negative emotions influence different objects of trust through information-seeking behavior during an emerging public health crisis.

The *psychological demand* (emotions) for information sufficiency serves as the primary driver behind the proactive search and methodological processing of risk information in the RISP model [48]. Research indicates that information-seeking and avoidance can be motivated by both positive and negative emotions [49-52]. Scholars revealed that people with positive emotions tend to have a holistic perception of seeking information more comprehensively during a pandemic [46].

Negative emotions, such as fear and anxiety, are common psychological reactions when facing a crisis such as the Middle East respiratory syndrome, H1N1 influenza, and H7N9 influenza [10,53,54]. In the early stage of the COVID-19 pandemic, the preventive measures imposed by the government exacerbated prevalent negative emotions such as frustration, hostility, and anger [55,56]. Previous research has examined the relationship between emotions and information use behavior, for instance, Bohner and Weinerth [57] demonstrated that while positive emotion reduces people's tendency to examine information, negative affect enhances it. Park et al [47] found that people's emotional reactions, such as worry and fear, were triggered by their perceived risk, and this exacerbated the perception of their own lack of information.

We suspect that negative emotions are positively associated with information engagement. Therefore, we formulated this hypothesis:

- H1: Negative emotions have a positive relationship with information engagement at the early stage of a pandemic.

The RISP model posits that affective responses and information insufficiency might enhance information-seeking and engaging behaviors that are related to one's trust [48,58]. Previous studies have investigated the impact of negative emotions, such as anger, sadness, and anxiety, on trust during public health crises [10,59,60]. A line of studies found that negative emotions decrease trust [61-63]. However, Zhang et al [64] found that negative emotions—particularly anger—could increase trust. Researchers have also examined the relationships between emotions, information seeking or engagement, and trust, particularly regarding trust in government. For example, Ahn et al [65] made cross-cultural comparisons and concluded that early in the COVID-19 pandemic in South Korea, the United States, and Singapore, trust in the government was negatively related to and reduced by negative emotions (anger, fear, sadness, and anxiety), and positively related to hope. In addition, these emotions acted as a mediator in the relationship between trust and information seeking. Erhardt et al [66] discovered that negative emotions have an impact on public trust in the government during times of crisis. For instance, anger places the blame for unfavorable circumstances on the government. Khalifa [67] reported that among the Bahrainis, trust in the government during the COVID-19 pandemic was positively connected with the use of news websites, but negatively associated with dependency on television and social media. On the basis of the aforementioned theoretical frameworks and findings, this research aims to dive into the relationships between negative emotions, information engagement, and trust in the different key objects during the early stage of the COVID-19 pandemic in China.

Because previous research has no consensus on the relationship between negative emotions, information engagement behavior, and trust, this study chose to propose the following 2 research questions (RQs):

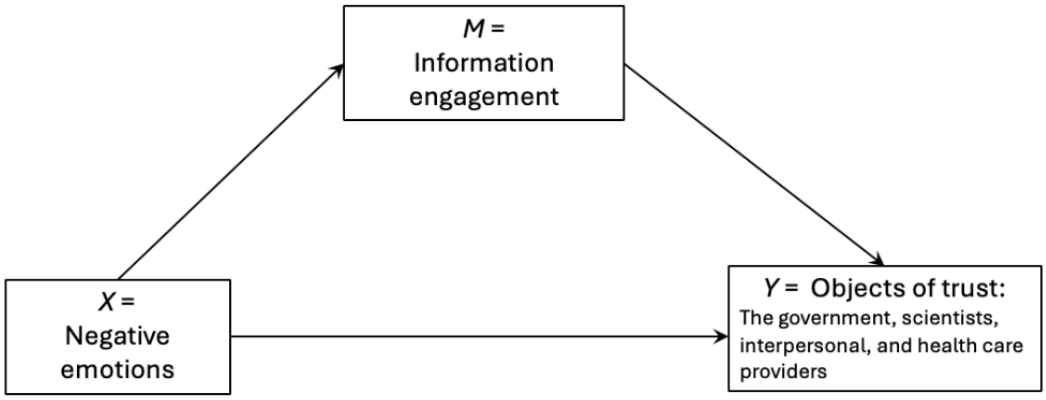
- RQ1: How are negative emotions associated with trust in different objects?
- RQ2: How does trust in different objects relate to negative emotions and information engagement?

In our data, we have 4 different but significant objects of trust: the government, scientists, others (interpersonal trust), and health care providers. How each is operationalized is explained in the Methods section.

To summarize the assumptions of the relationships between negative emotions, information seeking, and trust in various entities, Figure 1 is the conceptual model that theorizes the relationships between these 3 variables. This study examines the potential mediating roles of media use, drawing on the media practice model, which conceptualizes media engagement and its outcomes as part of a reciprocal, cyclical process rather than a straightforward linear effect [68,69]. Specifically, we suspected that information engagement mediates the association between negative emotions and trust. Therefore, we created 4 mediation models, one with each object of trust, to be tested.

XSL•FO
RenderX

Figure 1. The proposed model of the relationships between variables.



Methods

Data

An online survey was conducted during the early stage of the COVID-19 outbreak in China, from January 31, 2020, to February 9, 2020. The data collection was particularly valuable due to the limited access to national samples in China during the early stages of the public health crisis. Sojump, a professional Chinese internet survey company that is similar to SurveyMonkey, gathered the data. It first randomly selected 2840 people from its sample pool and then sent them an email invitation to participate. A total of 1656 people completed the survey, with a response rate of 58.3%. This sampling technique was a time-efficient way to investigate public emotions and opinions during the COVID-19 outbreak. Previous research conducted in China has taken this approach [70-72].

Participants who completed the survey (a questionnaire with 179 items) in fewer than 11 minutes [73] or failed the attention checks were excluded, ending with 1568 valid cases. A total of 31 provinces, municipalities, and autonomous territories in mainland China were included in the sample. With a mean age of 31.02 (SD 9.0), there were 789 (50.32%) male respondents and 779 (49.68%) female respondents. Furthermore, 18.11% (284/1568) of the participants had a monthly household income of ≤CN ¥5000; 31.12% (488/1568) of the participants earned between CN ¥5001 and CN ¥10,000; 21.17% (332/1568) of the participants earned between CN ¥10,001 and CN ¥15,000; 15.75% (247/1568) of the participants earned between CN ¥15,001 and CN ¥20,000; 12.4% (194/1568) of the participants earned between 20,001 and 50,000 CN ¥; and 1.47% (23/1568) of the participants earned >CN ¥50,000 CN ¥. On a 1-to-10-point scale of income, the mean is 6.29 (SD 1.96), which is in the category of CN ¥8001 to CN ¥10,000. Only 5.04% (79/1568) of the sample had completed high school or less, 16.07% (252/1568) were current or former technical school students, 69.07% (1083/1568) were college students or alumni, and 9.82% (154/1568) were postgraduates. On a 1 to 9-point scale of education, the mean was 6.76 (SD .92), which falls into the category of some or professional college.

Ethical Considerations

The instrument and data collection received approval from the Institutional Review Board of Zhejiang University (2020-056). Prior to completing the online survey, all participants provided voluntary informed consent. Each respondent received a compensation of CN¥ 12 (US $1.60) for their participation. The dataset was fully anonymized, with no personally identifiable information retained or linked to any participant.

Measurements

The dependent variables (*Y*) measure one's trust in four distinct entities: (1) the local government in terms of its response to the pandemic (Cronbach α=0.85); (2) scientists studying and battling the virus (Cronbach α=0.79); (3) interpersonal actors (family, relatives, friends, and strangers regarding the disclosure of their health conditions and travel histories; Cronbach α=0.66); and (4) the health care providers who treated COVID-19 infections (Cronbach α=0.75).

In the initial stage of a crisis, individuals often experience a constellation of negative emotions [47,74]. To capture the overall emotional response during the early stage of the public health crisis, the independent variable (*X*) is a combination of 5 negative emotions—fear, anxiety, anger, frustration, and hostility (Cronbach α=0.79). This approach follows prior research that conceptualizes emotional arousal as a general affective state influencing information processing and behavioral intention [10,53]. The mediator (*M*) is a scale of how often people engaged with different information-related activities in the context of COVID-19, such as posting an original message or liking, commenting on, or forwarding or sharing messages that they had encountered (Cronbach α=0.72).

Similar research has studied the same variables [75-78]. Control variables are typical demographics in mass and health communication research, including sex, age, income, and education [79]. Table 1 summarizes these variables. The PROCESS Macro (Model 4, version 4.1) proposed by Hayes [80] for SPSS was used to analyze the data, with 5000 bootstrap samples. Continuous variables were mean-centered before analysis. We created a mediation model for each of the 4 objects of trust. Only models with statistical significance are reported.

XSL•FO
RenderX

Table 1. A summary of dependent, independent, and control variables.

Variables and wording	Mean (SD)	Cronbach α
Dependent variable 1: trust in the government[a]		.85
How do you rate the performance of the local government of where you live? (average scale)	3.90 (0.70)	
1. I have confidence in the ability of local government departments in terms of controlling the pandemic.	4.02 (0.79)	
2. Relevant local government departments will fully consider the benefits of the people in the context of the pandemic.	3.85 (0.89)	
3. Relevant local government departments show fairness when dealing with the pandemic.	3.80 (0.91)	
4. Relevant local government departments are transparent when conveying information about the pandemic.	3.80 (1.0)	
5. Overall, relevant local government departments are trustworthy in their handling of the pandemic.	4.03 (0.84)	
Dependent variable 2: trust in scientists[a]		.79
How do you rate the scientists who are on the front line fighting COVID-19? (average scale)	4.35 (0.56)	
1. They deserve my trust.	4.36 (0.64)	
2. Their professional qualifications are trustworthy.	4.38 (0.70)	
3. They put people's health as a topic priority.	4.30 (0.80)	
4. I trust them.	4.37 (0.71)	
Dependent variable 3: trust–interpersonal[b]		.66
When you ask the following people about their recent travel histories and health conditions, how much do you trust their answers? (average scale)	3.63 (0.55)	
1. Family	4.53 (0.71)	
2. Relatives	3.80 (0.84)	
3. Friends	3.85 (0.75)	
4. Strangers	2.35 (0.81)	
Dependent variable 4: trust in health care providers[a]		.75
How much do you agree with the following statements? (average scale)	3.51 (0.75)	
1. The confirmation rates of COVID-19 in hospitals are low (reverse-coded).	3.91 (0.76)	
2. Hospitals can treat patients effectively because they are overcrowded and have a shortage of supplies (reverse-coded).	2.94 (1.11)	
3. Physicians do not have enough knowledge about COVID-19 and are prone to misdiagnosis (reverse-coded).	3.93 (0.95)	
4. Hospitals are prone to misdiagnosis because they do not have enough COVID-19 test kits (reverse-coded).	3.32 (1.19)	
5. Hospitals cannot effectively isolate COVID-19 patients, which will likely cause cross-contamination (reverse-coded).	3.45 (1.22)	
***X*: negative emotions**[a]		.79
The country is extremely concerned about the current COVID-19 pandemic. How much do you exhibit or feel each of the following emotions during this outbreak? (average scale)	3.73 (0.82)	
1. Fear	2.91 (1.03)	
2. Anxiety	2.97 (1.16)	
3. Anger	2.43 (1.27)	
4. Frustration	2.16 (1.03)	
5. Hostility	1.69 (0.96)	
***M*: engagement with information**[c]		.72

XSL•FO
RenderX

Variables and wording	Mean (SD)	Cronbach α
When you are exposed to online information related to the COVID-19 pandemic, how do you respond to the information? (average scale)	2.85 (0.77)	
1. Give a like	2.83 (2.20)	
2. Post original message	1.91 (1.07)	
3. Comment	2.68 (1.14)	
4. Forward or reshare	3.11 (1.18)	
5. Search further for related information	3.75 (1.03)	
Control 1		
Age (range 16 to 67 years)	31.02 (9.0)	—[d]
Control 2		
Sex (male: 50.3%; female: 49.7%); dummy-coded male	—	—
Control 3		
Education (1-9 scale; 1=no schooling; 9=doctorate)	6.76 (0.92)	—
Control 4		
Monthly income (1-10 scale; 1=no income; 10= ≥CN ¥50,001 [e])	6.29 (1.96)	—

[a]1=totally disagree, 2=disagree, 3=neither disagree nor agree, 4=agree, and 5=totally agree.

[b]1=totally trust, 7=totally distrust.

[c]1=none, 2=selfdom, 3=several times, 4=often, and 5=very frequently.

[d]Not available.

[e]CN ¥1=US $0.14.

Results

The first hypothesis posits that the more negative emotions one has, the more they are likely to engage with information. The mediation model shown in Table 2 indicates that there was a positive association (B=0.219, SE 0.023; *P*<.001) between negative emotions and information engagement. Therefore, H1 was supported.

RQ1 asks about the relationship between negative emotions and trust in different objects. Higher levels of negative emotions were linked to *more* trust in the government, scientists, and interpersonal trust, but *less* trust in health care providers.

RQ2 asks whether information engagement has a mediating effect on the path between negative emotions and trust in 4 different entities. The answer to RQ2 is that information engagement had a mediating effect between negative emotions and trust in the government and scientists, but *not* in health care providers or interpersonal trust.

The trust-in-the-government mediation model was supported, as Table 2 and Figure 2 demonstrate. The a_1-path from negative emotion to information engagement (*X* predicting *M*) was significant (B=0.219, SE 0.023; *P*<.001), suggesting that individuals experiencing more negative emotions such as fear, anger, and anxiety tended to be more information-engaging. There was also a significant and positive coefficient (B=0.074, SE 0.023; *P*=.001) for the b_1-path from information engagement to trust in the government (*M* predicting *Y*). Also notable was the c'_1 -path coefficient (X predicting Y), which reflects the direct effect of negative emotions on trust in the government (B=0.191, SE 0.022; *P*<.001). The total effect from *X* (negative emotions) to *Y* (trust in the government) was 0.207, SE 0.021; *P*<.001. The indirect effect from *X* (negative emotions) to *Y* (trust in the government) was 0.016, SE 0.01, 95% CI 0.01-0.027. Because 0.01 and 0.027 are both on the same side of zero, this indirect impact was validated. This set of findings suggests that an individual's trust in the government was positively associated with their level of negative emotions. In addition, experiencing *more* negative emotions encouraged people to engage with *more* information, which *raised* their level of trust in the government.

Figure 3 and the data in Table 2 show that the trust-in-the-scientists mediation model stood. Evidence of the connection between negative emotions and information engagement (*X* predicting *M*) is found in the a_2-path (B=0.219, SE 0.023; *P*<.001). The b_2-path from information engagement to trust in scientists (*M* predicting *Y*) had a significant and positive coefficient (B=0.054, SE 0.018; *P*=.003). Another significant relationship between negative emotions and trust in scientists was the c'_2-path (*X* predicting *Y*) coefficient (B=0.184, SE 0.017; *P*=.003). The total effect from *X* (negative emotions) to *Y* (trust in scientists) was 0.196; SE 0.017; *P*<.001. The indirect effect from *X* (negative emotions) to *Y* (trust in scientists) was 0.012; SE 0.01; 95% CI 0.004-0.02. Because 0.004 and 0.02 are both on the same side of zero, the indirect effect through mediation was confirmed. This implies that a person's degree of negative emotions rose with their degree of trust in scientists. Furthermore, feeling *more* negative emotions motivated people to engage with *more* information, hence *increasing* their level of trust in scientists.

XSL•FO
RenderX

Table 2. Trust in the government and scientists regression models generated by PROCESS Model 4.

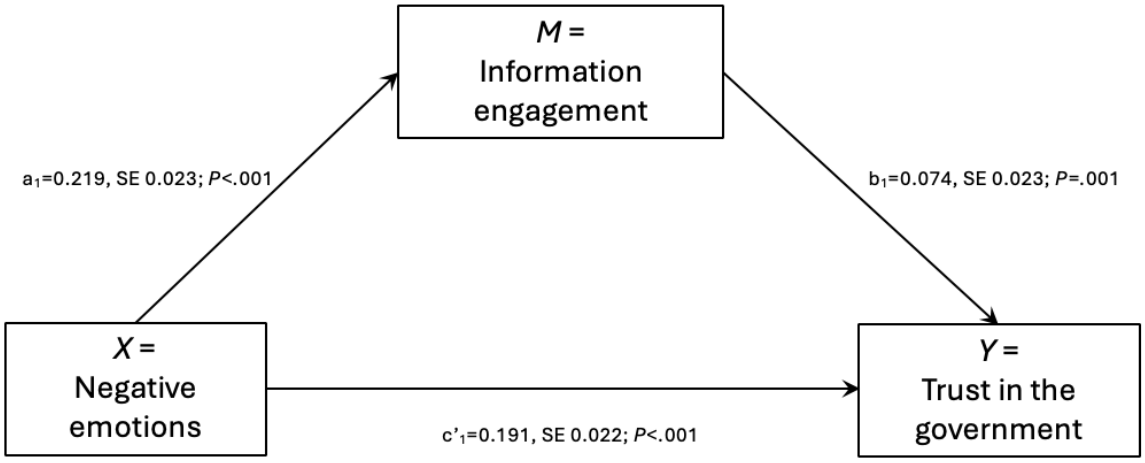
Variables	B (SE; 95% CI)	*P* value
Predicting information engagement[a]		
Negative emotions	0.219 (0.023; 0.174 to 0.265)	<.001
Age (years)	−0.006 (0.002; −0.010 to –0.001)	.02
Gender	0.006 (0.038; −0.069 to 0.080)	.88
Education	0.038 (0.022; −0.006 to 0.081)	.09
Income	0.045 (0.012; 0.024 to 0.066)	<.001
Predicting trust in the government[b]		
Negative emotions	0.191 (0.022; 0.148 to 0.233)	<.001
Information engagement	0.074 (0.023; 0.029 to 0.120)	.001
Age (years)	0.005 (0.002; 0.001 to 0.010)	.001
Gender	0.024 (0.035; −0.044 to 0.092)	.49
Education	−0.026 (0.020; −0.066 to 0.014)	.21
Income	0.015 (0.010; −0.005 to 0.035)	.13
Predicting trust in scientists[c]		
Negative emotions	0.184 (0.017; 0.150 to 0.218)	<.001
Information engagement	0.054 (0.018; 0.018 to 0.090)	.003
Age (years)	−0.002 (0.002; −0.005 to 0.001)	.26
Gender	−0.009 (0.027; −0.062 to 0.045)	.75
Education	−0.009 (0.016; −0.040 to 0.022)	.57
Income	−0.008 (0.008; −0.023 to 0.007)	.29

[a]R^2=0.070; $F_{5,1562}$=23.46.

[b]R^2=0.073; $F_{6,1561}$=20.41.

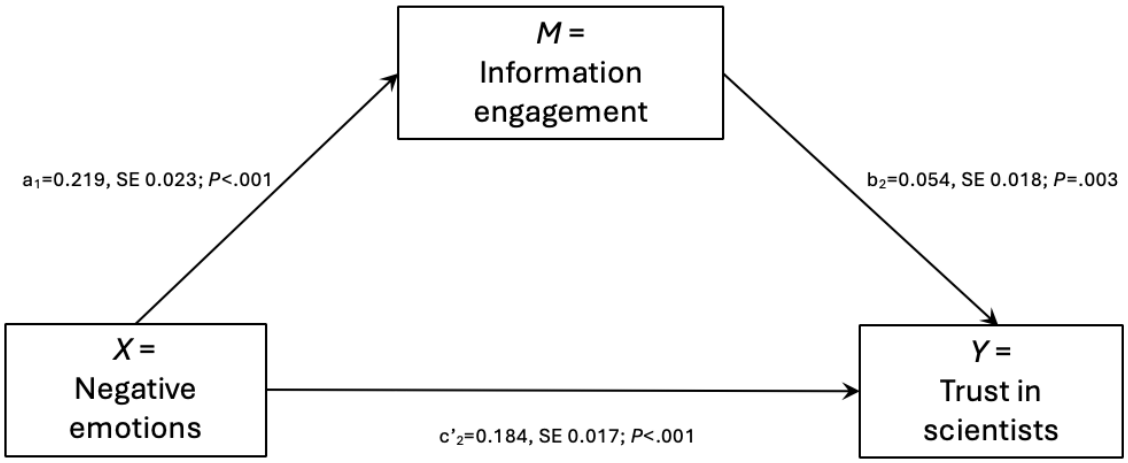
[c]R^2=0.089; $F_{6,1561}$=25.38.

Figure 2. The mediating role of information engagement in the relationship between negative emotions and government trust.



XSL•FO
RenderX

Figure 3. The mediating role of information engagement in the relationship between negative emotions and trust in scientists.



Discussion

Principal Findings

This research reveals that individuals' negative emotions are both directly and indirectly (through information engagement) positively associated with their trust in government and scientists. However, these negative emotions do not affect respondents' trust in others (relatives, friends, family, and strangers) or health care providers. These discrepancies with previous findings [65] call for further investigation into factors that influence public responses during health crises and risks, particularly in the initial stages of a public health emergency [47].

The Objects of Trust

Although negative emotions tend to encourage a higher level of engagement with information, echoing previous findings [46,57,65], the consequent effect on trust relies heavily on the specific entities involved. The fact that negative emotions and information engagement are positively associated with trust in the government and scientists, but have no connection with trust in others (interpersonal trust) or health care providers, underscores the necessity for a more intricate understanding of the interplay between the dynamics of trust.

As argued earlier, the objects of trust can be roughly divided into 2 camps: persons versus institutions or systems [14,24,30]. The analysis reveals that the negative emotions and information engagement are positively associated with trust toward *institutional* or *impersonal* entities in our models, namely the government and scientists who supposedly possessed more knowledge about how to combat the deadly virus. These 2 entities are not individual "everyday people" and they tend not to have direct contact with the general population. The government is associated with formal roles and responsibilities that impact the broader community yet may not have direct personal interactions with most individuals. Scientists were the authoritative heroes in a far way laboratory figuring out how to battle the virus. So both the government and scientists were *impersonal* to our survey participants, and they were the ones that our survey respondents likely looked up to and trusted.

In addition, trust in government in the Chinese context is multilayered and dynamic, often differing across levels (central vs local) and shaped by both political culture and past governance performance [81]. Previous studies suggest that local governments in China typically receive less trust than central authorities, particularly during uncertain times [82]. However, this study examines how local government trust and digital engagement interact during unique circumstances where negative emotions can drive information-seeking behavior that enhances—rather than weakens—institutional trust.

Conversely, findings suggest that during the early stage of the COVID-19 pandemic in China, individuals' negative emotions and information engagement were not associated with their trust in the types of persons that they had *direct and tangible contact* with in their daily lives. Family members and friends are within everyday social interactions and health care providers are the ones in the hospital who can be reached when necessary. COVID-19 is a contagious virus that spreads among people and interpersonal trust in China has been decreasing [83]. Any person could be a virus carrier or transmitter. People either did not know whether they had the virus or had an incentive to lie about their health conditions or travel histories so that they would not be quarantined. Therefore, other people were less likely to be trusted in this context, leading to an insignificant outcome in our analysis. In addition, during the initial stage of this pandemic, health care providers likely knew about the virus and how to treat infections as little as ordinary citizens. Therefore, trust in health care providers is not a significant factor in our findings.

The Effects of Information Engagement

In the initial stage of the COVID-19 pandemic in China, the pervasive sense of fear, anger, and anxiety likely triggered a heightened state of negative emotions among the populace. Faced with a novel and rapidly evolving crisis, individuals probably sought out information that could provide them with a sense of security, reassurance, and guidance. Selective and

XSL•FO
RenderX

avoidance exposure originated from the theory of cognitive dissonance, which holds that people experience discomfort or dissonance when they are exposed to facts or situations that contradict their personal beliefs [84,85]. Selective exposure—a psychological strategy used to reduce feelings of dissonance—involves deliberately avoiding information that contradicts one's views and seeking out information that supports existing beliefs.

When the public is faced with uncertainty and lacks the competence to make decisions, trust in scientists plays a crucial role in science-related decisions [27,86]. Following the COVID-19 pandemic outbreak, particularly after January 20, 2020, when Zhong Nanshan, a highly regarded Chinese expert was sent to Wuhan by the National Health Commission to investigate the situation and confirm the human-to-human transmission of SARS-CoV-2, a unique form of collaboration between the government and representative scientists emerged [87]. The government created a communication matrix for scientists, and scientists acted as vital microphones for the government to inform the public of important information [87,88]. The communication matrix accompanied by the highly centralized media system made Chinese news media the major information source. The media focused on the government's and scientists' responses to the pandemic, highlighting preemptive steps and decisive acts, which could increase public confidence in their capabilities [89]. Consequently, the government and scientists, as primary sources of information and guidance during the crisis, gained prominence in the media and public discourse [11,40,90].

Individuals experiencing heightened negative emotions were likely to selectively expose themselves to information disseminated by trusted governmental and scientific authorities. The lack of direct personal involvement with these authoritative figures might have contributed to a more idealized perception of their capabilities and intentions. In contrast, during the early stage of the pandemic, the public knew that hospitals were overwhelmed, leading to the likely perception that health care providers could not handle the situation adequately.

Sociocultural Factors

While earlier studies, notably Ahn et al [65], suggested a *negative* correlation between negative emotions and trust, our study revealed a *positive* relationship, namely negative emotions *fostering* an increase in trust toward governmental authorities and scientists.

Cultural frameworks contribute to explaining variations in belief and behavior patterns across nations. China is collectivist with a greater power distance [91-93]. In other words, Chinese culture emphasizes collective harmony and societal order. Therefore, Chinese people—in comparison with citizens in individualist cultures—would respect and trust the government as well as authority figures such as scientists more [40,94], particularly in uncertain times. To preserve societal stability and cohesiveness, the cultural tendency of deference to authority and hierarchy may encourage people to believe in the government and scientists, particularly in the face of negative emotions.

Moreover, a greater sense of dependence on government-led initiatives and outlets occurred among the Chinese during times of crisis due to its highly centralized governmental structure, and a tradition of information control [95,96]. At the early stage of a public health crisis, even in the face of unpleasant emotional experiences, citizens may view the government and scientific authorities as important sources of direction and assistance, particularly when facing difficulties.

Implications

This study offers important theoretical contributions to the fields of risk communication and crisis management. The RISP model [13,48], which highlights the role of emotions in shaping trust and information engagement during crises, provides the foundation for this research, which extends the literature by examining how negative emotions differentially influence trust toward various entities, including the government, scientists, health care providers, and interpersonal networks. By demonstrating that negative emotions primarily fostered trust in governmental and scientific institutions but not in health care providers or social relations, the findings underscore the need for a more nuanced understanding of trust formation processes. Moreover, this study emphasizes the importance of incorporating sociocultural factors into analyses of emotion-trust dynamics. The findings contradict previous studies [65] showing that revealed negative emotions and trust in the government are typically negatively associated in the face of a crisis. This research calls for a more nuanced, context-sensitive approach in theorizing about emotion, information engagement, and trust during public health crises.

Practically, the findings of this study offer valuable insights for future risk management and communication strategies during public health emergencies. Understanding that negative emotions can selectively reinforce trust in certain institutions highlights the need for targeted communication efforts that recognize the differential credibility assigned to various sources [12]. Risk communicators and policy makers should design emotion-sensitive strategies that not only acknowledge public emotional states but also actively guide information engagement toward credible institutions to strengthen societal resilience [1]. Furthermore, by recognizing the influence of sociocultural contexts on trust dynamics, practitioners can better tailor their interventions to the specific cultural and institutional environments in which crises unfold [65]. These lessons provide a critical foundation for developing more effective, culturally adaptive risk communication frameworks that can better manage uncertainty, foster trust, and promote collective action in the face of future crises.

Limitations and Suggestions for Future Research

This study has several limitations that should be acknowledged. First, the sample is not fully representative of the broader population, and the data collection period is relatively brief. With 69.07% (1083/1568) of the respondents being current university students or graduates, and 9.82% (154/1568) being postgraduates, highly educated people are overrepresented. This skewness is a common attribute observed in online nonprobability survey samples, potentially limiting the generalizability of the findings [97]. Nevertheless, given the

XSL•FO
RenderX

unique circumstances of the crisis's initial outbreak and infeasibility to access the representative sample, this method represented the most effective approach for the data collection. Second, while the RISP model includes key constructs such as perceived risk, this variable was not incorporated into our model. Instead, this study focused on how negative emotions influence trust in different entities, offering a theoretical extension by emphasizing the object-specific nature of trust. In addition, because this research gathered data in China, its findings may have limited generalizability to other populations.

Notwithstanding these limitations, we believe this research has yielded valuable knowledge. We have learned about how people deal with emotions, information engagement, and trust at the earliest stage of a national or global crisis. This timing provided a unique opportunity to examine the Chinese public's trust dynamics and emotional responses during crisis management at an early stage when uncertainty was at its peak and society faced unprecedented unknowns.

Future research should use a more representative survey sample and ask questions about trust across different levels of the government as well as the use of additional types of information sources. Including open-ended questions would help reveal why respondents trust or distrust specific entities. In addition, researchers should investigate distinct dimensions of interpersonal trust—between close ties and strangers—to develop a more nuanced understanding of its nature.

Conclusions

This study revisited the early stage of the COVID-19 pandemic in China to examine the complex interplay between negative emotions, information engagement, and trust in different actors. The findings revealed that while negative emotions directly and indirectly enhanced trust in the government and scientists through increased information engagement, they did not significantly impact trust in health care providers or interpersonal relationships. This study has contributed to a deeper understanding of the Chinese context, in comparison with other countries and cultures. The observed positive relationship between negative emotions, information engagement, and trust in the government and authoritative scientists likely signifies a distinctive pattern specific to the Chinese sociocultural context, diverging from the trends identified in previous studies conducted in different national settings, such as in the United States, Korea, Singapore [65], and Bahrain [67]. In addition, there is an observed difference between scientists and health care providers as trustees in the face of a pandemic. These noteworthy distinctions underscore the influence of unique cultural and societal norms when the chips are down in shaping individuals' perceptions and responses, highlighting the importance of contextual factors in shaping trust dynamics within diverse sociocultural landscapes.

This study offers valuable lessons for designing future strategies for public health crisis and risk communication. Recognizing the selective nature of trust formation and tailoring communication efforts to emotional and cultural contexts can enhance the effectiveness of crisis response. As the world continues to face emerging public health threats, this type of retrospective analysis provides vital insights for developing more resilient and effective public health communication systems.

Acknowledgments

This work was supported by a major project of the National Social Science Fund of China "A study on Media System and Social Trust" (19ZDA325).

Conflicts of Interest

None declared.

References

1. Coombs WT. Ongoing Crisis Communication: Planning, Managing, and Responding. Thousand Oaks, CA: SAGE Publications; Dec 18, 2018.
2. Flor LS, Friedman J, Spencer CN, Cagney J, Arrieta A, Herbert ME, et al. Quantifying the effects of the COVID-19 pandemic on gender equality on health, social, and economic indicators: a comprehensive review of data from March, 2020, to September, 2021. Lancet 2022 Jun 25;399(10344):2381-2397 [FREE Full text] [doi: 10.1016/S0140-6736(22)00008-3] [Medline: 35247311]
3. Liu Z, Tu J, Lee TT, Wei L. Does trust affect behavior in a public health crisis? Testing an extended theory of planned behavior model with trust. J Conting Crisis Manage 2025 Apr 25;33(2):e70051. [doi: 10.1111/1468-5973.70051]
4. de Macedo Steffen G, Lacerda DP, Richter C, de Souza Gomes RF. Leading adaptive risk context: contingent plans of Covid-19 at universities. Int J Disaster Risk Reduct 2025 Apr 01;120:105385. [doi: 10.1016/j.ijdrr.2025.105385]
5. Li Y, Zeng Y, Liu G, Lu D, Yang H, Ying Z, et al. Public awareness, emotional reactions and human mobility in response to the COVID-19 outbreak in China – a population-based ecological study. Psychol Med 2020 Sep 25;52(9):1793-1800. [doi: 10.1017/s003329172000375x]
6. Su Y, Wu P, Li S, Xue J, Zhu T. Public emotion responses during COVID-19 in China on social media: an observational study. Hum Behav Emerg Technol 2020 Dec 10;3(1):127-136. [doi: 10.1002/hbe2.239]

XSL•FO
RenderX

7. Xie X, Zang Z, Ponzoa JM. The information impact of network media, the psychological reaction to the COVID-19 pandemic, and online knowledge acquisition: evidence from Chinese college students. J Innov Knowl 2020;5(4):297-305. [doi: 10.1016/j.jik.2020.10.005]
8. Zhao Y, Xu S, Wang L, Huang Y, Xu Y, Xu Y, et al. Concerns about information regarding COVID-19 on the internet: cross-sectional study. J Med Internet Res 2020 Nov 09;22(11):e20487 [FREE Full text] [doi: 10.2196/20487] [Medline: 33095740]
9. Dong W, Tao J, Xia X, Ye L, Xu H, Jiang P, et al. Public emotions and rumors spread during the COVID-19 epidemic in China: web-based correlation study. J Med Internet Res 2020 Nov 25;22(11):e21933 [FREE Full text] [doi: 10.2196/21933] [Medline: 33112757]
10. Wei L, Huang Q. Retrospecting digital media use, negative emotions, and trust gaps during the COVID-19 pandemic in China: cross-sectional web-based survey. J Med Internet Res 2024 Jul 10;26:e49422 [FREE Full text] [doi: 10.2196/49422] [Medline: 38986127]
11. Guo D, Liu S, Sun Y. Who can help me? Citizens' help-seeking on Weibo during the Shanghai lockdown. Disaster Med Public Health Prep 2025 Jan 03;18:e329. [doi: 10.1017/dmp.2024.325] [Medline: 39749776]
12. Boin A, McConnell A, Hart P. Governing the Pandemic: The Politics of Navigating a Mega-Crisis. Cham, Switzerland: Springer; 2021.
13. Shah Z, Wei L. Interpersonal risk communication matters more than media risk communication in its impact on individuals' trust and preventive behaviors during COVID-19. Int J Disaster Risk Reduct 2022 Nov;82:103369 [FREE Full text] [doi: 10.1016/j.ijdrr.2022.103369] [Medline: 36267112]
14. Hardin R. The street-level epistemology of trust. Politics Soc 1993 Dec;21(4):505-529. [doi: 10.1177/0032329293021004006]
15. Robbins BG. What is trust? A multidisciplinary review, critique, and synthesis. Sociol Compass 2016 Oct 05;10(10):972-986. [doi: 10.1111/soc4.12391]
16. Rotter JB. Generalized expectancies for interpersonal trust. Am Psychol 1971;26(5):443-452. [doi: 10.1037/h0031464]
17. Tsfati Y, Cappella JN. Do people watch what they do not trust?: exploring the association between news media skepticism and exposure. Commun Res 2003 Oct;30(5):504-529. [doi: 10.1177/0093650203253371]
18. Han R, Xu J, Pan D. How media exposure, media trust, and media bias perception influence public evaluation of COVID-19 pandemic in international metropolises. Int J Environ Res Public Health 2022 Mar 25;19(7):3942 [FREE Full text] [doi: 10.3390/ijerph19073942] [Medline: 35409623]
19. Zhao E, Wu Q, Crimmins EM, Ailshire JA. Media trust and infection mitigating behaviours during the COVID-19 pandemic in the USA. BMJ Glob Health 2020 Oct;5(10):e003323 [FREE Full text] [doi: 10.1136/bmjgh-2020-003323] [Medline: 33037063]
20. Blöbaum B. Key factors in the process of trust. On the analysis of trust under digital conditions. In: Trust and Communication in a Digitized World. Cham, Switzerland: Springer; Feb 25, 2016.
21. van der Meer TW, Ouattara E. Putting 'political' back in political trust: an IRT test of the unidimensionality and cross-national equivalence of political trust measures. Qual Quant 2019 Aug 14;53:2983-3002. [doi: 10.1007/s11135-019-00913-6]
22. Zmerli S, Newton K. Objects of political and social trust: scales and hierarchies. In: Zmerli S, van der Meer TW, editors. Handbook on Political Trust. Cheltenham, UK: Edward Elgar Publishing; Jan 27, 2017.
23. Alon I, Bar-Tal D. The Role of Trust in Conflict Resolution: The Israeli-Palestinian Case and Beyond. Berlin, Germany: Springer International Publishing; 2017.
24. Levi M. A state of trust. In: Braithwaite V, Levi M, editors. Trust and Governance. New York, NY: SAGE Publications; 1998:77-101.
25. Safari A, Barzoki AS, Heidari Aqagoli P. Exploring the antecedents and consequences of impersonal trust. Int J Organ Anal 2020 Oct 19;28(6):1149-1173. [doi: 10.1108/ijoa-08-2019-1850]
26. Flanagan CA, Stout M. Developmental patterns of social trust between early and late adolescence: age and school climate effects. J Res Adolesc 2010 Sep 01;20(3):748-773 [FREE Full text] [doi: 10.1111/j.1532-7795.2010.00658.x] [Medline: 20936077]
27. Yuan H, Long Q, Huang G, Huang L, Luo S. Different roles of interpersonal trust and institutional trust in COVID-19 pandemic control. Soc Sci Med 2022 Jan;293:114677 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114677] [Medline: 35101260]
28. Raven J, Wurie H, Witter S. Health workers' experiences of coping with the Ebola epidemic in Sierra Leone's health system: a qualitative study. BMC Health Serv Res 2018 Apr 05;18(1):251 [FREE Full text] [doi: 10.1186/s12913-018-3072-3] [Medline: 29622025]
29. Stefaniak A, Wohl MJ, Elgar FJ. Commentary on "Different roles of interpersonal trust and institutional trust in COVID-19 pandemic control". Soc Sci Med 2022 Apr;299:114765 [FREE Full text] [doi: 10.1016/j.socscimed.2022.114765] [Medline: 35125233]
30. Domański H, Pokropek A. The relation between interpersonal and institutional trust in European countries: which came first? Polish Sociol Rev 2021;213(1):87-102 [FREE Full text] [doi: 10.26412/psr213.05]
31. Letki N. Investigating the roots of civic morality: trust, social capital, and institutional performance. Polit Behav 2006 Oct 27;28(4):305-325. [doi: 10.1007/s11109-006-9013-6]

XSL•FO
RenderX

32. Offe K. How can we trust our fellow citizens? In: Warren ME, editor. Democracy and Trust. Cambridge, UK: Cambridge University Press; 1999:42-87.
33. Luhmann N, Davis H, Raffan J, Rooney K, King M, Morgner C. Trust and Power. Hoboken, NJ: John Wiley & Sons; 2017.
34. Vanhala M, Puumalainen K, Blomqvist K. Impersonal trust: the development of the construct and the scale. Pers Rev 2011 Jun 07;40(4):485-513. [doi: 10.1108/00483481111133354]
35. Salmon DA, Dudley MZ, Glanz JM, Omer SB. Vaccine hesitancy: causes, consequences, and a call to action. Vaccine 2015 Nov 27;33 Suppl 4:D66-D71. [doi: 10.1016/j.vaccine.2015.09.035] [Medline: 26615171]
36. Schmelz K. Enforcement may crowd out voluntary support for COVID-19 policies, especially where trust in government is weak and in a liberal society. Proc Natl Acad Sci U S A 2021 Jan 05;118(1):e2016385118 [FREE Full text] [doi: 10.1073/pnas.2016385118] [Medline: 33443149]
37. Gilson L. Trust and the development of health care as a social institution. Soc Sci Med 2003 Apr;56(7):1453-1468. [doi: 10.1016/s0277-9536(02)00142-9] [Medline: 12614697]
38. Torpe L, Lolle H. Identifying social trust in cross-country analysis: do we really measure the same? Soc Indic Res 2010 Sep 18;103(3):481-500. [doi: 10.1007/s11205-010-9713-5]
39. Wei L, Yan C. Social quality of China: indicators, reality, and problems. Int J Soc Quality 2018;8(2):78-102. [doi: 10.3167/IJSQ.2018.080205]
40. Liu SN, Chang TK. One disaster, three institutional responses: legitimation crisis and competing discourses in China. Journal Stud 2016 Jun 09;19(3):392-414. [doi: 10.1080/1461670x.2016.1190666]
41. Bursztyn L, Rao A, Roth CP, Yanagizawa-Drott DH. Misinformation during a pandemic. National Bureau of Economic Research. 2020. URL: https://www.nber.org/system/files/working_papers/w27417/w27417.pdf [accessed 2025-08-08]
42. Yang F, Huang Z. Health communication and trust in institutions during the COVID-19 lockdown in China's urban communities. Urban Governance 2021 Nov;1(1):17-22. [doi: 10.1016/j.ugj.2021.10.001]
43. Algan Y, Cohen D, Davoine E, Foucault M, Stantcheva S. Trust in scientists in times of pandemic: panel evidence from 12 countries. Proc Natl Acad Sci U S A 2021 Oct 05;118(40):e2108576118 [FREE Full text] [doi: 10.1073/pnas.2108576118] [Medline: 34580225]
44. Funk C, Hefferon M, Kennedy B, Johnson C. Trust and mistrust in Americans' views of scientific experts. Pew Research Center. 2019 Aug 2. URL: https://www.pewresearch.org/science/2019/08/02/trust-and-mistrust-in-americans-views-of-scientific-experts/ [accessed 2025-08-08]
45. Leonard MB, Pursley DM, Robinson LA, Abman SH, Davis JM. The importance of trustworthiness: lessons from the COVID-19 pandemic. Pediatr Res 2022 Feb;91(3):482-485 [FREE Full text] [doi: 10.1038/s41390-021-01866-z] [Medline: 34853429]
46. Huang J, Ren W, Wang S, Zhou Y, Yang Y. Positive emotion and media dependence: measuring risk information seeking and perception in the COVID-19 pandemic prevention. Inquiry 2023 Mar 15;60:469580231159747 [FREE Full text] [doi: 10.1177/00469580231159747] [Medline: 36919995]
47. Park T, Ju I, Ohs JE, Hinsley A, Muzumdar J. Information seeking during the COVID-19 pandemic: application of the risk information seeking and processing model. Res Social Adm Pharm 2023 Jun;19(6):956-964. [doi: 10.1016/j.sapharm.2023.03.010] [Medline: 37024346]
48. Griffin RJ, Dunwoody S, Neuwirth K. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ Res 1999 Feb;80(2 Pt 2):S230-S245. [doi: 10.1006/enrs.1998.3940] [Medline: 10092438]
49. de los Santos TM, Nabi RL. Emotionally charged: exploring the role of emotion in online news information seeking and processing. J Broadcast Electron Media 2019 Mar 18;63(1):39-58. [doi: 10.1080/08838151.2019.1566861]
50. Khan SS. Developing positive emotion through affective design for interactive information seeking. In: Proceedings of the Seventh BCS-IRSG Symposium on Future Directions in Information Access. 2017 Presented at: FDIA 2017; September 5, 2017; Barcelona, Spain. [doi: 10.14236/ewic/fdia2017.0]
51. Li J, Zheng H. Online information seeking and disease prevention intent during COVID-19 outbreak. Journal Mass Commun Q 2020 Sep 28;99(1):69-88. [doi: 10.1177/1077699020961518]
52. Lyons J, Sokhey A. Emotion, motivation, and social information seeking about politics. Polit Commun 2014 Apr 16;31(2):237-258. [doi: 10.1080/10584609.2013.828138]
53. Chen X, Huang B, Lin W. The effect of interpersonal relationship and epidemic attention on negative emotion among medical students: the mediating role of social satisfaction. BMC Psychiatry 2023 Aug 21;23(1):610 [FREE Full text] [doi: 10.1186/s12888-023-05113-z] [Medline: 37605114]
54. Hao R, Liu Y, Shen W, Zhao R, Jiang B, Song H, et al. Surveillance of emerging infectious diseases for biosecurity. Sci China Life Sci 2022 Aug 04;65(8):1504-1516 [FREE Full text] [doi: 10.1007/s11427-021-2071-x] [Medline: 35287183]
55. Han S, Sun T, Sun Y, Gao X. Public psychological crisis reports: an investigation during the remission of COVID-19 in China. Chin Manag Stud 2023 Jan 2;17(1):112-129. [doi: 10.1108/cms-08-2021-0351]
56. Dubey S, Biswas P, Ghosh R, Chatterjee S, Dubey MJ, Chatterjee S, et al. Psychosocial impact of COVID-19. Diabetes Metab Syndr 2020;14(5):779-788 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.035] [Medline: 32526627]

XSL•FO
RenderX

57. Bohner G, Weinerth T. Negative affect can increase or decrease message scrutiny: the affect interpretation hypothesis. Pers Soc Psychol Bull 2001 Nov 01;27(11):1417-1428. [doi: 10.1177/01461672012711003]
58. So J, Kuang K, Cho H. Information seeking upon exposure to risk messages: predictors, outcomes, and mediating roles of health information seeking. Commun Res 2016 Nov 24;46(5):663-687. [doi: 10.1177/0093650216679536]
59. Folkman S, Moskowitz JT, Ozer EM, Park CL. Positive meaningful events and coping in the context of HIV/AIDS. In: Gottlieb BH, editor. Coping with Chronic Stress. Boston, MA: Springer; 1997:293-314.
60. Vemprala N, Bhatt P, Valecha R, Rao HR. Emotions during the COVID-19 crisis: a health versus economy analysis of public responses. Am Behav Sci 2021 Oct 25;65(14):1972-1989. [doi: 10.1177/00027642211050901]
61. Rodriguez-Sanchez C, Schuitema G, Claudy M, Sancho-Esper F. How trust and emotions influence policy acceptance: the case of the Irish water charges. Br J Soc Psychol 2018 Jul;57(3):610-629. [doi: 10.1111/bjso.12242] [Medline: 29392740]
62. Smith N, Leiserowitz A. The role of emotion in global warming policy support and opposition. Risk Anal 2014 May;34(5):937-948 [FREE Full text] [doi: 10.1111/risa.12140] [Medline: 24219420]
63. Yuan Y, Yang S, Jiang X, Sun X, Lin Y, Liu Z, et al. Trust in government buffers the negative effect of rumor exposure on people's emotions. Curr Psychol 2022 Jul 30:1-14 (forthcoming) [FREE Full text] [doi: 10.1007/s12144-022-03508-x] [Medline: 35967504]
64. Zhang K, Goetz T, Chen F, Sverdlik A. Angry women are more trusting: the differential effects of perceived social distance on trust behavior. Front Psychol 2021 Jul 21;12:591312 [FREE Full text] [doi: 10.3389/fpsyg.2021.591312] [Medline: 34366949]
65. Ahn J, Kim HK, Kahlor LA, Atkinson L, Noh GY. The impact of emotion and government trust on individuals' risk information seeking and avoidance during the COVID-19 pandemic: a cross-country comparison. J Health Commun 2021 Oct 03;26(10):728-741. [doi: 10.1080/10810730.2021.1999348] [Medline: 34779340]
66. Erhardt J, Freitag M, Filsinger M, Wamsler S. The emotional foundations of political support: how fear and anger affect trust in the government in times of the covid-19 pandemic. Schweiz Z Polit 2021 Jun;27(2):339-352. [doi: 10.1111/spsr.12462] [Medline: 35923361]
67. Khalifa H. COVID-19 pandemic and trust in government: the case of Bahrain. Int J Manag 2020 Nov;11(11):329-336. [doi: 10.34218/IJM.11.11.2020.032]
68. Bobkowski PS, Shafer A, Ortiz RR. Sexual intensity of adolescents' online self-presentations: joint contribution of identity, media consumption, and extraversion. Comput Hum Behav 2016 May;58:64-74. [doi: 10.1016/j.chb.2015.12.009]
69. Shafer A, Bobkowski P, Brown JD. Sexual media practice: how adolescents select, engage with, and are affected by sexual media. In: The Oxford Handbook of Media Psychology. Oxford, UK: Oxford University Press; Jan 28, 2013.
70. Chen D, Cheng CY, Urpelainen J. Support for renewable energy in China: a survey experiment with internet users. J Clean Prod 2016 Jan 20;112(5):3750-3758. [doi: 10.1016/j.jclepro.2015.08.109]
71. Huang Q. Understanding public perceptions of genetically modified organisms in China: the role that heuristics play during digital media exposure. Chin J Commun 2019 Oct 08;13(3):293-311. [doi: 10.1080/17544750.2019.1673453]
72. Lien CH, Cao Y. Examining WeChat users' motivations, trust, attitudes, and positive word-of-mouth: evidence from China. Comput Hum Behav 2014 Dec;41:104-111. [doi: 10.1016/j.chb.2014.08.013]
73. Sharma H. How short or long should be a questionnaire for any research? Researchers dilemma in deciding the appropriate questionnaire length. Saudi J Anaesth 2022;16(1):65-68 [FREE Full text] [doi: 10.4103/sja.sja_163_21] [Medline: 35261591]
74. Yu W, Shen F. Does fact-checking habit promote COVID-19 knowledge during the pandemic? Evidence from China. Public Health 2021 Jul;196:85-90 [FREE Full text] [doi: 10.1016/j.puhe.2021.05.005] [Medline: 34166857]
75. Karnowski V, Kümpel AS, Leonhard L, Leiner DJ. From incidental news exposure to news engagement. How perceptions of the news post and news usage patterns influence engagement with news articles encountered on Facebook. Comput Hum Behav 2017 Nov;76:42-50. [doi: 10.1016/j.chb.2017.06.041]
76. Kim HK, Niederdeppe J. The role of emotional response during an H1N1 influenza pandemic on a college campus. J Public Relations Res 2013 Jan 10;25(1):30-50. [doi: 10.1080/1062726x.2013.739100]
77. Mayer RC, Davis JH. The effect of the performance appraisal system on trust for management: a field quasi-experiment. J Appl Psychol 1999;84(1):123-136. [doi: 10.1037//0021-9010.84.1.123]
78. Oeldorf-Hirsch A. The role of engagement in learning from active and incidental news exposure on social media. Mass Commun Soc 2017 Oct 20;21(2):225-247. [doi: 10.1080/15205436.2017.1384022]
79. Li Y, Wang X. Seeking health information on social media: a perspective of trust, self-determination, and social support. J Organ End User Comput 2018;30(1). [doi: 10.4018/JOEUC.2018010101]
80. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: Guilford Publications; 2013.
81. Zhai Y. Government policy performance and central–local political trust in China. J Public Policy 2022 Aug 09;42(4):782-801. [doi: 10.1017/s0143814x22000162]
82. Chen D. Local distrust and regime support: sources and effects of political trust in China. Pol Res Q 2017 Feb 06;70(2):314-326. [doi: 10.1177/1065912917691360]
83. Yang Z, Xin Z. Income inequality and interpersonal trust in China. Asian J Soc Psychol 2019 Dec 23;23(3):253-263. [doi: 10.1111/ajsp.12399]

XSL•FO
RenderX

84. Festinger L. A Theory of Cognitive Dissonance. Redwood City, CA: Stanford University Press; 1957.
85. Harmon-Jones E, Mills J. An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In: Harmon-Jones E, editor. Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology. Washington, DC: American Psychological Association; 2019.
86. Hamilton LC, Hartter J, Saito K. Trust in scientists on climate change and vaccines. Sage Open 2015 Aug 25;5(3). [doi: 10.1177/2158244015602752]
87. Liu P, Zhong X, Yu S. Striking a balance between science and politics: understanding the risk-based policy-making process during the outbreak of COVID-19 epidemic in China. J Chin Gov 2020 Apr 24;5(2):198-212. [doi: 10.1080/23812346.2020.1745412]
88. Ren F. A reconsideration on telling good stories about Chinese scientists and shaping the exemplary image of scientists amid the anti-pandemic campaign. Cultur Sci 2021 Feb 17;3(4):245-259. [doi: 10.1177/2096608321994002]
89. Rieger MO, Wang M. Trust in government actions during the COVID-19 crisis. Soc Indic Res 2022;159(3):967-989 [FREE Full text] [doi: 10.1007/s11205-021-02772-x] [Medline: 34456450]
90. Zhang T, Yu L. The relationship between government information supply and public information demand in the early stage of COVID-19 in China-an empirical analysis. Healthcare (Basel) 2021 Dec 31;10(1):77 [FREE Full text] [doi: 10.3390/healthcare10010077] [Medline: 35052242]
91. Hofstede G. Culture′s Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. Thousand Oaks, CA: SAGE Publications; 2001.
92. House RJ, Hanges PJ, Javidan M, Dorfman PW, Gupta V. Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies. Thousand Oaks, CA: SAGE Publications; 2004.
93. Sztompka P. Trust in the moral space. In: Sasaki M, editor. Trust in Contemporary Society. Leiden, The Netherlands: Brill; 2019:31-40.
94. Yang J, Dong C, Chen Y. Government's economic performance fosters trust in government in China: assessing the moderating effect of respect for authority. Soc Indic Res 2021 Jan 03;154:545-558. [doi: 10.1007/s11205-020-02553-y]
95. He AJ, Shi Y, Liu H. Crisis governance, Chinese style: distinctive features of China's response to the Covid-19 pandemic. Policy Des Pract 2020 Jul 30;3(3):242-258. [doi: 10.1080/25741292.2020.1799911]
96. Zhong K, Liu Y, Christensen T. Crisis coordination in centralized regimes: explaining China's strategy for combatting the COVID-19 pandemic. Int Public Manag J 2022 May 12;25(7):1131-1150. [doi: 10.1080/10967494.2022.2073411]
97. Lehdonvirta V, Oksanen A, Räsänen P, Blank G. Social media, web, and panel surveys: using non‐probability samples in social and policy research. Policy Internet 2020 Apr 29;13(1):134-155. [doi: 10.1002/poi3.238]

Abbreviations

RISP: risk information seeking and processing
RQ: research question

Edited by E Mensah; submitted 20.05.25; peer-reviewed by D Guo, SN Liu; comments to author 16.06.25; revised version received 26.06.25; accepted 01.08.25; published 25.08.25.

Please cite as:
Liu Z, Tu J, Lee TT, Wei L
Reconsidering Trust and Information Engagement and Unpacking the Role of Emotion in Public Responses During the Early Stage of a Public Health Crisis in China: Web-Based Survey Study
Online J Public Health Inform 2025;17:e77790
URL: https://ojphi.jmir.org/2025/1/e77790
doi:10.2196/77790
PMID:

©Zhiming Liu, Jiawei Tu, Tien-Tsung Lee, Lu Wei. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 25.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evidence-Based Mental Health at Scale: Benchmarking Retrospective Cohort Study of a Digital Employee Benefits Program for Depression and Anxiety

Emily J Ward[1], PhD; Matt Hawrilenko[1], PhD; Millard Brown[1], MD; Adam M Chekroud[1,2], PhD

[1]Spring Health, New York, NY, United States

[2]Department of Psychiatry, Yale School of Medicine, Yale University, New Haven, United States

Corresponding Author:
Emily J Ward, PhD
Spring Health
60 Madison Ave
New York, NY, 10010
United States
Phone: 1 917 517 8407
Email: eward@springhealth.com

Abstract

Background: Depression and anxiety affect millions worldwide; yet, many people face barriers to timely and effective mental health care, underscoring the need for scalable, high-quality interventions.

Objective: This study aimed to evaluate the clinical effectiveness and quality of a centralized, employer-sponsored mental health program in treating depression and anxiety during a period of rapid growth in access to mental health care.

Methods: This retrospective cohort study included participants using a digital mental health benefit (Spring Health), sponsored by 589 US employers from 2021 to 2024. Participants had access to therapists, psychiatrists, and care navigators. Primary measures were clinical effectiveness (treatment duration, Patient Health Questionnaire 9-item depression scale, Generalized Anxiety Disorder 7-item Scale) and clinical outcomes (reliable change, recovery, and remission). Outcomes were benchmarked to meta-analytic results of evidence-based therapy.

Results: A total of 52,929 adult participants started therapy from among 6868 providers during the study period, scored positive for depression or anxiety, and had at least one mental health assessment before and during treatment. Depression symptoms decreased with each log-day in treatment, resulting in a total reduction of 6.91 (95% CI –6.84 to –6.98) points at one-week post treatment, corresponding to a large effect size (d=1.61; 95% CI 1.60-1.63), significantly greater than the meta-analytic pre-post benchmark for psychotherapy (effect size difference=0.13, z=15.6, P<.001). Anxiety symptoms also decreased, resulting in a total reduction of 6.01 points (95% CI –5.95 to –6.08), corresponding to a large effect size (d=1.82; 95% CI 1.80 to 1.84), significantly greater than the meta-analytic benchmarks (effect size difference=0.64, z=61.9, P<.001). White participants and participants of color had similar outcomes. Logistic regression showed 92.3% (95% CI 92.0% to 92.5%) of participants' symptoms (in depression or anxiety) reliably improved, and 61.7% (95% CI 61.1% to 62.4%) achieved remission by 1-week posttreatment.

Conclusions: Among a large and diverse sample, using a digital mental health benefit with a centralized system of care produces clinical outcomes in depression and anxiety significantly greater than what is typically observed through meta-analyses of psychotherapy. By using data to monitor, incentivize, and improve quality of care, the clinical outcomes outperform or equal benchmarks as a growing number of individuals across race, gender, and age access mental health care.

(Online J Public Health Inform 2025;17:e72999) doi:10.2196/72999

KEYWORDS

depression and anxiety; measurement-based care; employer-sponsored mental health; clinical outcome evaluation; digital mental health benefit

XSL•FO
RenderX

Introduction

Depression and anxiety affect over 280 million people worldwide [1,2], with over 9% of Americans having depression and 18% having anxiety [3]. Although many barriers to mental health care continue to prevent people from getting the type of treatment and quality of treatment they need [4-6], mental health has become a priority across public and private sectors. As employee-benefit packages providing mental health care have proliferated [7], the focus is increasingly on optimizing care: delivering the right care at the right time to ensure positive clinical outcomes [8,9] and financial return on investment through reduced health care spending [8,10] and increased workplace productivity [8,11].

However, simply offering mental health services does not guarantee quality care, and scaled programs without methods to promote and monitor evidence-based treatment can undermine their potential impact [12]. Measurement-based care [13,14]—regularly assessing symptom change and clinical progress—can help address this challenge by ensuring treatment is continuously evaluated for effectiveness. At scale, large data sets offer unprecedented opportunities to evaluate outcomes across diverse populations [15], track provider outcomes, and refine patient-provider matching so that each patient's clinical profile and treatment preferences are respected. But any mental health solution at scale must first ensure an adequate number of qualified mental health professionals to meet demand and to have positive clinical outcomes.

One promising strategy for promoting accessible, high-quality care is to build broad provider networks that offer a variety of empirically supported treatments. On one hand, a broad provider network could increase treatment accessibility and patient satisfaction, potentially leading to better outcomes. On the other hand, this flexibility necessitates evaluating treatment outcomes. A recent review of digital mental health solutions demonstrated that programs that offer a variety of treatment options show strong positive outcomes, such as the percentage of patients seeing statistically significant improvements in depression, ranging from mid-60% to low-80% [16]. The outcomes of these programs should also be compared with established clinical benchmarks, such as those produced by specific treatment modalities such as cognitive behavioral therapy (CBT) (eg, Cohen *d* or equivalent metrics of 0.63 for depression [17] and 0.51 for anxiety [18]) to ensure the sustainability of high quality outcomes within diverse service settings and among heterogeneous populations.

This retrospective cohort study aimed to evaluate the therapeutic effectiveness of an employer-sponsored mental health program serving 53,000 patients with depression and anxiety through a network of more than 6800 providers. Leveraging the scale of these patient-provider pairs, we aimed to benchmark the program's effectiveness overall, explore outcomes among different demographic subgroups, and determine how effect sizes changed as the program scaled. By capitalizing on extensive measurement data, we provide insights into how large-scale programs can preserve high-quality care through timely access, tailored provider matching, and rigorous monitoring of clinical outcomes.

Methods

Ethical Considerations

This retrospective cohort study was approved by the Yale Institutional Review Board (protocol ID: 2000029276) and classified as not involving research with human participants. Informed consent was not required, as the data were anonymized prior to analysis. The research adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for observational studies. While some authors had access to identifiable information during and after data collection, all analyses were performed using deidentified data. The privacy and confidentiality of research participants' data and identity was maintained. Participants did not receive compensation.

Program Design

Data came from an employer-sponsored digital mental health benefit (Spring Health; Spring Care Inc). The program incorporates evidence-based components to increase access to and use of mental health care as follows.

After completing an initial assessment (see Assessments), participants received a personalized care plan and were able to schedule appointments. Care recommendations ranged from using a library of digital self-help modules (eg, mindfulness exercises, CBT skill-builders) or coaching for low to low-moderate risk concerns. For moderate to high risk concerns, participants could engage in unlimited, no-cost video appointments with care navigators who assisted with finding appropriate care, or could browse the network of clinical care providers and self-schedule appointments directly. Regardless of risk level, all participants had access to free or low-cost access to psychotherapy (with additional sessions offered as an in-network benefit covered by their health plan) and medication management through video or in-person sessions. Providers listed their clinical specialties and modalities on their profiles, with the predominant modality being CBT. Across these different care options, the platform delivers care through a centralized data system in which all provider notes, patient data, assessment results, patient scheduling, and payment are integrated into one system. In the centralized system of care, average provider availability was one day nationwide. Patients could directly book their chosen provider's appointment. Measurement-based care was implemented, and incentive structures rewarded and retained top-performing providers.

Therapists held master or doctoral-level licenses, and all medication providers were either physicians or psychiatric nurse practitioners. All clinicians had at least 3 years of postsupervision experience before joining the network. A total of 6868 providers provided care to participants in the study.

To lower the financial barriers to accessing care, participants could book an unlimited number of free appointments with a care navigator and had access to a varying number of employer-sponsored sessions with a therapist or medication provider (median 12, IQR 8-14). Additional sessions with the

XSL•FO
RenderX

same providers could be continued with copays and deductibles according to behavioral health plan selection and coverage, often as in-network coverage with platform providers for over 50% of cases.

Setting

Treatment was primarily delivered through videoconferencing, and assessments were collected using the program platform. The study period was from January 1, 2021, to June 1, 2024, and only the first epoch of care (no more than 6 months could elapse between therapy sessions) was considered. Appointment and assessment data were collected up to 6 months after the first therapy appointment. Final assessments that occurred beyond 1 SD of the average interval from the final appointment were excluded prior to analyses (ie, within about 35 [–18 days before end of treatment + 53] days of the final appointment). The final study size was determined by how many participants met the inclusion criteria (see Participants).

Participants

Participants were employees and dependents eligible for the benefit during the study period. Participants were included if they took an assessment and started treatment within the study period, if they were located in the United States, were older than 18 years, had a baseline assessment within 30 days of starting treatment or before the second appointment if they screened positive for depression or anxiety (Patient Health Questionnaire 9-item Scale [PHQ-9]≥10 or Generalized Anxiety Disorder 7-item Scale [GAD-7]≥10), and if they had an additional assessment after treatment began. This criterion aimed to capture a broader real-world population engaging with the service while still ensuring a pretreatment baseline and at least one follow-up for outcome measurement. Participants must have remained eligible for the benefit for at least 6 months after starting, across which period data were collected. Participants did not receive a monetary reward.

Assessments

The program included a digital mental health assessment tool, which enabled a proactive symptom monitoring system to support measurement-based care. Within the digital platform, individuals completed a series of questionnaires to identify common mental health difficulties (such as stress, anxiety, sleeping, eating, or relationship issues). They completed the PHQ-9 [19,20] for depression and additional self-report questionnaires based on the mental health issues identified by the participant. For those with anxiety, this included the GAD-7 [21]. The cut-off for a positive depression screening was PHQ-9≥10; for a positive anxiety screening, it was GAD-7≥10. Participants were reminded by email every 2 to 4 weeks while in care to complete follow-up assessments, but all assessments were optional, and participants were not required to complete an assessment at the end of treatment.

Measures

Depression and Anxiety

Depression symptoms were measured with the PHQ-9 (range 0-27), consisting of 9 items assessing the frequency of a range of depression symptoms. Anxiety symptoms were measured with the GAD-7 (range, 0-21), consisting of 7 items assessing the frequency of a range of anxiety symptoms. Both PHQ-9 and GAD-7 items are scored 0-3 (not at all to nearly every day). PHQ-9 and GAD-7 scores were outcome variables for assessing symptom change (continuous outcomes), reliable improvement [22] (5-point decrease in the PHQ-9 [19] and 4-point decrease in the GAD-7), and reliable improvement with recovery (both reliable improvement and ending in the subclinical range [20], corresponding to a PHQ-9 score <10 or a GAD-7 score of <10) and remission (ending with PHQ-9 and GAD-7 scores <5). Depression and anxiety symptoms were modeled separately.

Factors Associated With Clinical Improvement

The primary outcome was change in clinical symptoms over time (days in treatment, measured from the start of therapy [t=0], estimated at 1-week posttreatment, corresponding to the average duration of care across all patients plus one week [t=95.4 average duration + 7 days], and at the end of 6 months [t=180 days]). Demographics (age, race, and gender) were collected from employer census files or when individuals enrolled in the program. The program did not compel individuals to disclose personal information, such as their race or gender, and those choosing to share such information did so through free-response, thereby self-endorsing their own identities, which were categorized for data analysis or marked as not available ("Missing" in Table 1). No attempts were made to "predict" missing demographic information using any identifying personal information.

XSL•FO
RenderX

Table 1. Descriptive characteristics of the study sample.

Characteristics	Overall (N=52,929)
Race, n (%)	
White	9940 (61.2)
Asian	1166 (7.2)
Hispanic or Latino	1840 (11.3)
Two or more races/ethnicities	456 (2.8)
Black or African American	2737 (16.8)
American Indian or Alaska Native	111 (0.7)
Total	16,250 (30.7% of total)
Missing	36,679 (69.3% of total)
Gender, n (%)	
Man	2436 (24.3)
Woman	7341 (73.3)
Non-Binary	233 (2.3)
Total	10,010 (18.9% of total)
Missing	42,919 (81.1% of total)
Age group (years), n (%)	
18 to 30	13,751 (26.0)
30 to 50	30,117 (56.9)
50 to 65	8215 (15.5)
>65	846 (1.6)
Total	52,929 (100% of total)
Age (years)	
Mean (SD)	37.9 (11.2)
Median (Min-Max)	36.0 (18.0-88.0)
Baseline PHQ-9[a] (positive screening)	
Mean (SD)	15.4 (4.29)
Median (Min-Max)	15.0 (10.0-27.0)
N (%)	42,630 (80.5)
Baseline GAD-7[b] (positive screening)	
Mean (SD)	14.5 (3.31)
Median (Min-Max)	14.0 (10.0-21.0)
N (%)	39,860 (75.3)

[a]PHQ-9: Patient Health Questionnaire 9-item Scale.

[b]GAD-7: Generalized Anxiety Disorder 7-item Scale.

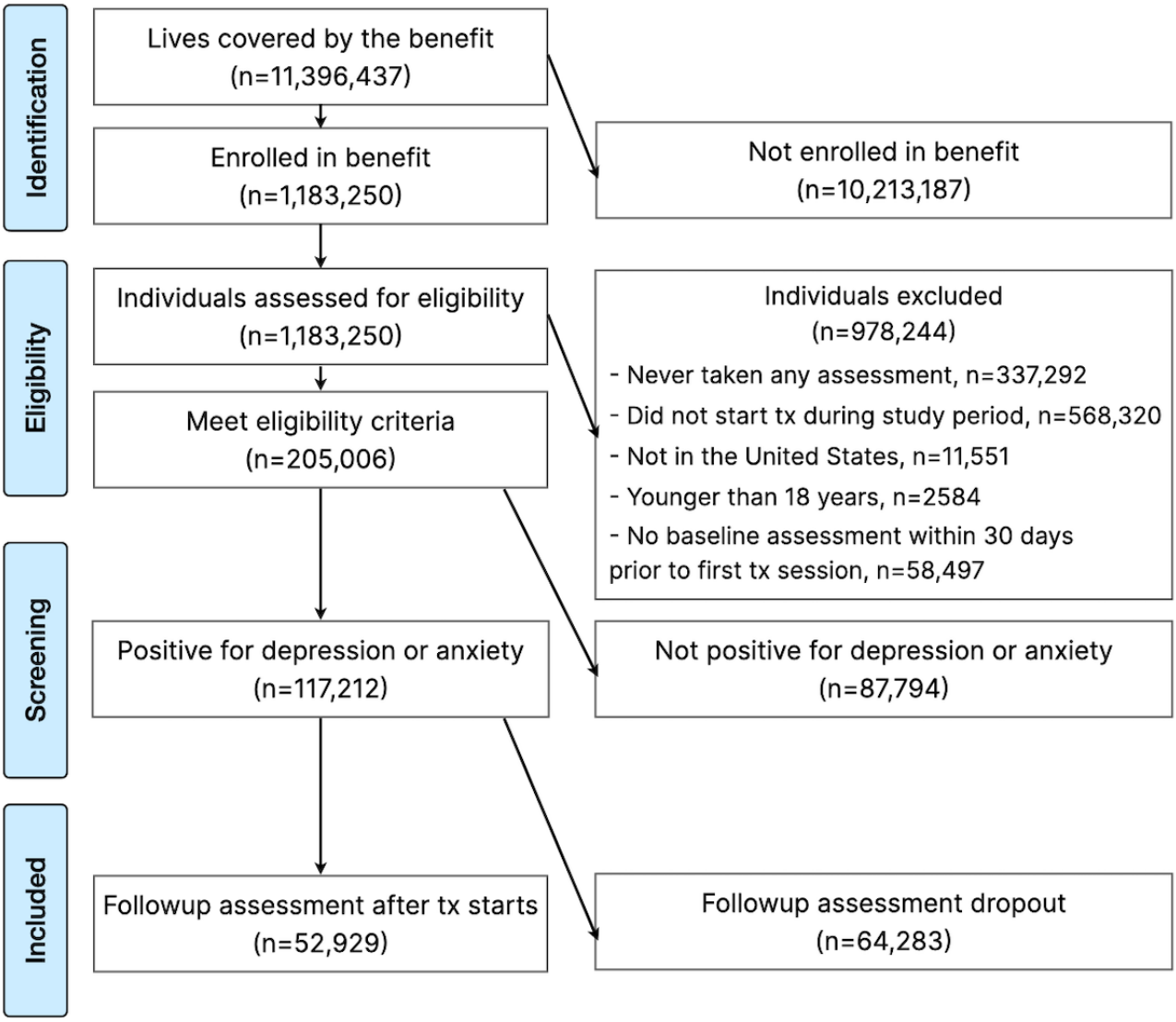
Missing Data

This study had 2 primary types of missing data. The first type of missing data was assessment drop-out, where patients did not complete any follow-up assessment after starting care, excluding 55% (n=64,283) of potential patients (Figure 1). The impact of this missingness was examined through 2 complementary approaches: (1) evaluating how treatment dosage (ie, number of therapy sessions) differed between those with and without follow-up assessments and (2) reweighting the data according to the treatment dosage distribution of those without follow-up assessments. The second type of missing data was demographic data, because neither race nor gender was required by the platform. To ensure our final sample was representative of our treatment-eligible population (those who enrolled in the benefit, n=1,183,250), we compared the racial and gender proportions using a chi-square test. Subgroup analyses are presented in accordance with best practice for equity-focused research [23,24], although due to the extent of missingness of

XSL•FO
RenderX

demographic data, these analyses should be considered exploratory.

Figure 1. Diagram depicting patient participant flow through the study from identification to study inclusion.



Meta-Analytic Benchmarks

Benchmarking single-study findings against meta-analytic estimates is standard practice in psychotherapy and health-services research, and can help improve the quality of interventions [25]. Effect sizes obtained from randomized control trials or meta-analyses have been used to evaluate the effectiveness of natural-setting psychotherapy [26] to large behavioral health agencies [27]. Furthermore, selecting an appropriate benchmark helps rule out improvement driven by extraneous factors or reversion to the mean in naturalistic samples.

In this study, the treatment effect sizes were compared with two types of meta-analytic benchmarks: (1) uncontrolled prepost effect sizes for psychotherapy and (2) the prepost effect size for treatment as usual (TAU). The uncontrolled prepost comparison can show the typical magnitude of within-group improvement that occurs over the course of psychotherapy, while the prepost TAU comparison can show whether an intervention has additional value above what is usually done in routine care (vs waitlist controlled trials often overestimate the effect sizes of psychotherapies [28]). There is heterogeneity across settings, and the meta-analytic effect size was derived from treatments spanning cognitive-behavioral, behavioral activation, interpersonal, third-wave, group, individual, and guided self-help formats. Furthermore, these studies comprised different study designs: most had 2 conditions, a subset had more than 2; about half used a cut-off score, and the rest used a diagnosis; participants came from community recruitment, clinical recruitment, or another method. While each individual study was testing a specific intervention, the meta-analysis summarizes the effect of therapy across all these different designs and interventions. But there are no significant differences between major categories of TAU when compared with psychotherapy conditions in randomized trials [29].

Data Analyses

We used mixed-effects regression models to estimate clinical outcomes. This approach is the widely accepted, gold-standard methodological approach to analyze outcomes—especially binary outcomes—with attrition and irregularly spaced measurements over time [30-32].

XSL•FO
RenderX

To address repeated observations for the same participants (such as multiple assessments over time), we used a mixed-effects model with repeated observations nested within participants as a random intercept. An identity link was used for continuous outcomes (depression and anxiety) and a logit link for categorical outcomes (reliable change, recovery, and remission). Symptom change, modeled using log-days since the start of treatment, was the main effect of interest. To determine which factors were associated with symptom change, the set of measures described above (age, gender, and race) was included as covariates. To account for the effect of provider on outcomes, provider was included as a random intercept; to investigate different trajectories of improvement at the participant and provider levels, a random slope term (ie, log-days in treatment) was included for both participant and provider.

Total improvement from the start of treatment to the desired endpoint (such as end of treatment or 180 days) was calculated using the delta method [33]. The delta method multiplies the rate of change (ie, the regression coefficient for log-days in treatment) by the desired duration to produce total improvement and standard errors from a function that combines model parameters. It avoids using the last observation carried forward, which has been shown to introduce bias [34,35].

Reliable improvement, recovery, and remission rates were predicted separately using logistic models with log-days since start of treatment as the main effect. Regression coefficients from these models were converted to probabilities to estimate the percentage of participants who showed reliable improvement, recovery, or remission from depression and anxiety symptoms within a modeled margin of error (ie, 95% CIs).

The treatment effect size was quantified using Cohen *d*, calculated by dividing the total treatment improvement in the measure of interest (eg, symptom-point decrease in PHQ-9) by the pre-test SD (eg, SD of PHQ-9 scores at baseline) to avoid a potential impact of care on the posttest SD [36]. There was a correlation (*r*) of 0.36 between PHQ-9 scores at baseline and at treatment duration, and a correlation (*r*) of 0.27 for GAD-7 scores. Effect sizes were categorized using established thresholds of small (*d* <0.50), medium (*d* <0.8), and large (*d* >0.8).

The treatment effect size was compared with meta-analytic benchmarks in two ways: (1) the overall prepost treatment effect was compared with uncontrolled prepost effect sizes for psychotherapy (depression: uncontrolled effect size=1.48 [28] and anxiety: uncontrolled effect size=1.18 [37]) and (2) to adjust for expected improvement in the course of usual care, the prepost effect size for TAU (depression: TAU effect size=0.64 [28] and anxiety: TAU effect size=0.59 [37]) was subtracted from the overall prepost treatment effect. The TAU-adjusted effect size was then compared with benchmark effect sizes for psychotherapy in randomized controlled trials (depression: effect size=0.63 [28] and anxiety: effect size=0.51 [37]) with TAU as the control group.

For example, an overall prepost treatment effect size of 1.50 would be compared with the uncontrolled effect size of 1.48; the TAU-adjusted effect size of 0.87 (1.50-0.64) would be compared with the RCT effect size of 0.63. All effect sizes were compared using a 1-sample z test. Using this TAU-adjustment is a means of approximating the true effect size of the program intervention in the absence of a control group.

To assess changes in effect size as a function of program scale, the difference between treatment effect size and benchmark values was predicted by the number of lives covered (in millions) using a linear regression.

All statistical tests were 2-sided with statistical significance set at α levels of .05. All analyses were conducted in R (version 2024.04.2; Posit, PBC) [38].

Results

Participant and Provider Characteristics

A total of 11,396,437 people were covered by the benefit during the study period, and 1,183,250 (10%) people enrolled (Figure 1). These individuals were assessed for eligibility, and 205,006 (17%) people met all eligibility criteria. Of these, 117,212 (57%) people screened positive for depression or anxiety. Finally, 52,929 (45%) people had an additional assessment after treatment began. This follow-up assessment was taken on average 34 (SD 36) days after the start of treatment, 14 (SD 21) days from the most recent session, and 64 (SD 56) days before the final session of treatment. The final assessment was taken on average 68 (SD 54) days after the start of treatment, 19 (SD 24) days from the most recent session, and 27 (SD 41) days before the final session of treatment.

Among the 52,929 participants who met the inclusion criteria (Table 1), their ages ranged from 18 to 88 (mean 37.9, SD 11.2) years. A total of 7341 of 10,010 (73.3%) participants with known gender information were female, and 9940 of 16,250 (61.2%) participants with known race information were White. This final sample of participants reflected 589 different employers. In addition, 42,630 participants screened positive for depression (scores ≥10 points) and 39,860 screened positive for anxiety (scores≥10 points), although not all provided race or gender. Whereas, 29,561 participants screened positive for both depression and anxiety, corresponding to a 55.9% comorbidity rate, similar to the general population [39-41]. Among the participants screening positive, baseline scores were moderate to severe for both PHQ-9 (mean 15.4, SD 4.3) and GAD-7 (mean 14.5, SD 3.3). Participants received care from 6868 providers.

Overall Clinical Outcomes

During the study period, the average number of business days to the next available appointment was 1.2 (SD 0.4 days. For example, if a patient searched for an appointment on Monday, the next available appointment they could book would be sometime on Tuesday. The average number of therapy sessions was 6.6 (SD 4.7), and the average number of medication management sessions was 2.7 (SD 1.8; 19.7% of participants attended at least one medication management session). Participants spent approximately 3 months in treatment (95.4, SD 58.5 days), with one-week posttreatment defined as 102 days. The average number of total assessments was 3.3 (which

XSL•FO
RenderX

included 1 baseline assessment and 2.3 assessments taken after therapy began).

Overall, PHQ-9 scores decreased with each log-day in treatment (*b*=–1.49, SE 0.01; *P*<.001) (Table 2), resulting in a total reduction of 6.91 (95% CI –6.84 to –6.98) points (Figure 2A) at one-week post-average-treatment duration, corresponding to a large effect size (*d*=1.61; 95% CI 1.60-1.63), significantly greater than the uncontrolled prepost effect size for psychotherapy [28] (effect size difference=0.13, *z*=15.6, *P*<.001). When adjusted for the expected effect of TAU, the effect size remained large (*d*=0.97; 95% CI 0.96-0.99) and significantly greater than meta-analytic benchmarks (effect size difference=0.34, *z*=40.6, *P*<.001).

GAD-7 scores also decreased with each log-day in treatment, (*b*=–1.30, SE 0.01; *P*<.001; Table 2), resulting in a total reduction of 6.01 (95% CI –5.95 to –6.08) points (Figure 2B), corresponding to a large effect size (*d*=1.82; 95% CI 1.80-1.84), significantly greater than the uncontrolled prepost effect size for psychotherapy (effect size difference=0.64, *z*=61.9, *P*<.001). When adjusted for the expected effect of TAU, the effect size remained large (*d*=1.23; 95% CI 1.21-1.25) and significantly greater than meta-analytic benchmarks (effect size difference=0.72, *z*=69.7, *P*<.001).

By the time of their final assessment (average of 68 days), 89.0% (n=47,131) of participants reliably improved or recovered as measured directly by change in their assessment scores. By 1-week posttreatment (102 days), results of mixed-effects logistic regression models (Figure 3) indicated that 92.3% of participants reliably improved or recovered (95% CI 92.0%-92.5%) and 61.7% (95% CI 61.1%-62.4%) achieved remission.

Table 2. Regression estimates of overall improvement in depression and anxiety symptoms.

	Depression (PHQ-9[a] score)	Anxiety (GAD-7[b] score)
Intercept		
Estimates	15.27	14.39
SE	0.02	0.02
P value	<.001[c]	<.001[c]
Time in treatment (log-days)		
Estimates	–1.49	–1.30
SE	0.01	0.01
P value	<.001[c]	<.001[c]
Random effects		
σ^2	12.94	10.79
τ_{00}		
Member_id	7.30	2.28
Provider_id	0.29	0.16
τ_{11}		
Member_id.time in tx	0.57	0.45
Provider_id.time in tx	0.03	0.03
ρ_{01}		
Member_id	0.22	0.94
Provider_id	–0.25	–0.28
Intraclass correlation	0.56	0.54
N		
Member_id	42,630	39,860
Provider_id	5797	5671
Observations	137,607	128,966
Marginal R^2 / Conditional R^2	0.225/0.658	0.217/0.637

[a]PHQ-9: Patient Health Questionnaire 9-item Scale.

[b]GAD-7: Generalized Anxiety Disorder 7-item Scale.

[c]Significant values (*P*<.05).

XSL•FO
RenderX

Figure 2. Change in symptoms of (A) depression, as measured using the PHQ-9, and (B) anxiety, as measured using the GAD-7. Overall symptom change is modeled by log-day in treatment and plotted in black on both panels. Each individual trajectory is plotted as an overlay and color-coded according to the probability of achieving either reliable change or recovery. GAD-7: Generalized Anxiety Disorder 7-item Scale; PHQ-9: Patient Health Questionnaire 9-item Scale.

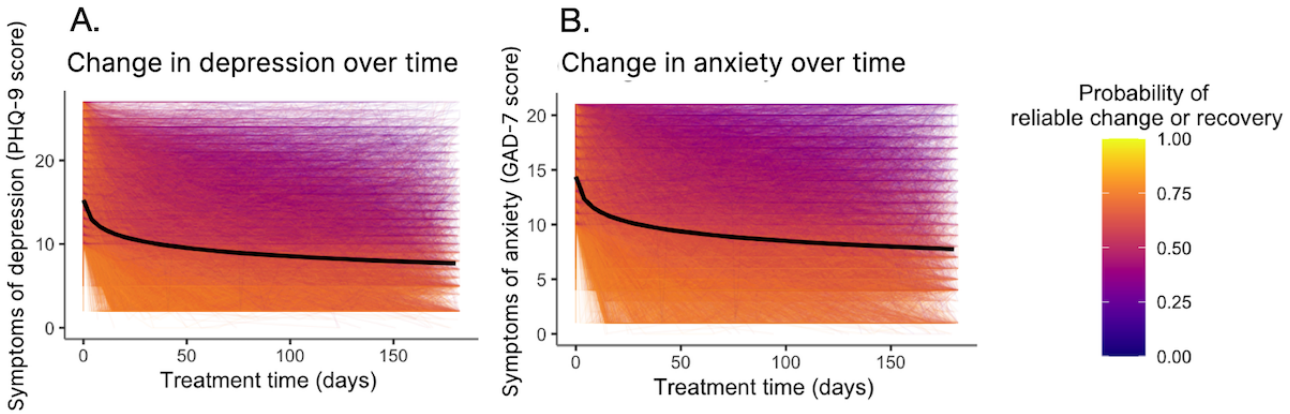
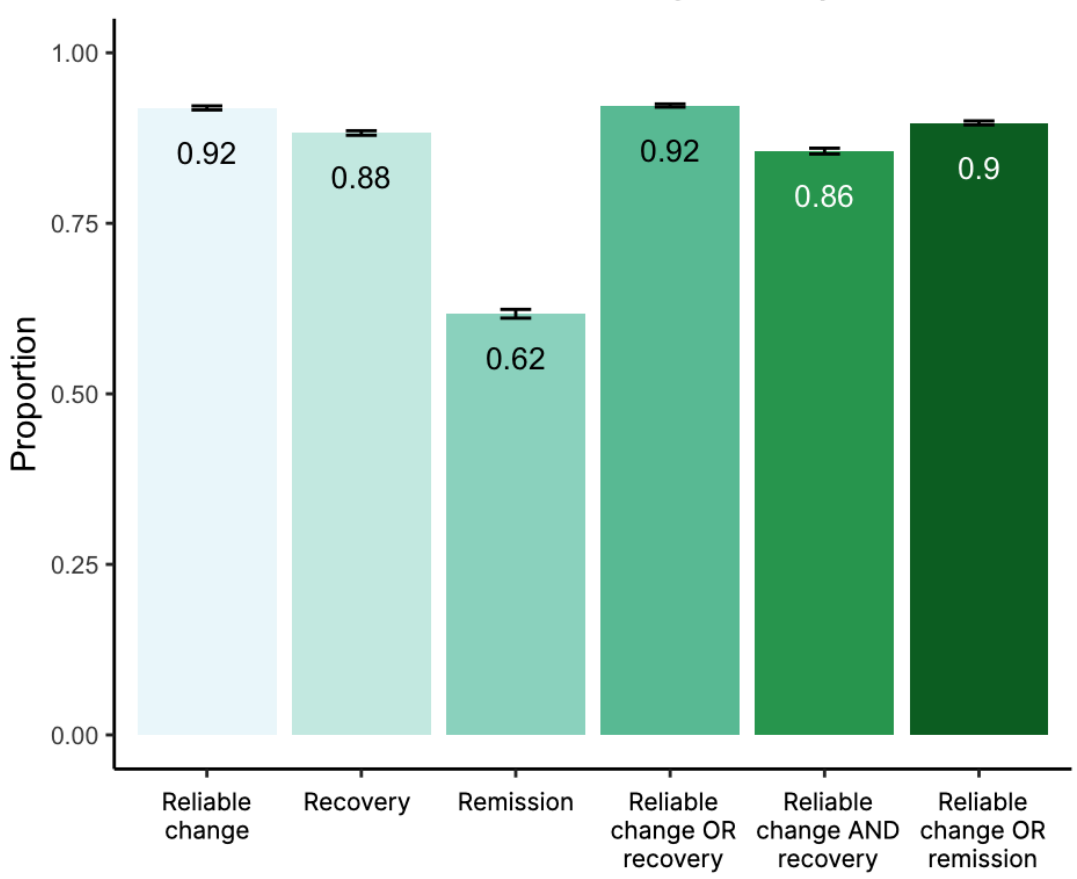


Figure 3. Clinical outcome results from a mixed-effects logistic regression for depression and anxiety. The predicted proportion of participants achieving reliable change (reduction of 5 points on Patient Health Questionnaire 9-item Scale [PHQ-9] or 4 points on Generalized Anxiety Disorder 7-item Scale [GAD-7]), recovery (scores of less than 10 points on PHQ-9 or less than 8 on GAD-7), or remission (scores of less than 5 points on either PHQ-9 or GAD-7) is plotted, along with the predicted proportion achieving combinations of clinical outcomes.



Follow-Up Assessment Sensitivity Analyses

Assessment drop-out—where patients did not complete any follow-up assessment after starting care—was a substantial source of missing data, excluding 55% of potential patients (Figure 1). Several sensitivity analyses were conducted to evaluate how treatment dosage (ie, number of therapy sessions) differed between those with and without follow-up assessments and to reweigh the data according to the treatment dosage distribution of those without follow-up assessments.

Overall, the baseline PHQ-9 and GAD-7 were similar between those who had a follow-up assessment (ie, those in the final sample) and those who did not. However, those who lacked a follow-up assessment attended far fewer total therapy sessions (4.0 sessions vs 6.6 sessions). In fact, 27% of those lacking a follow-up assessment attended only a single session (vs 6%

XSL•FO
RenderX

among those with over 2 assessments). To adjust our estimates, we included session count into our outcomes model, using maximum likelihood estimation to obtain an unbiased estimate of overall outcomes even when there are systematically fewer follow-up observations for the group of participants with fewer therapy sessions [42]. We re-estimated the PHQ-9 and GAD-7 models including all participants and added the centered session count as an interaction and found that while people who attended more sessions improved slightly more slowly per unit time (time in tx by number of sessions interaction: b=0.01, 95% CI 0.01-0.01, P<.001 for both PHQ-9 and GAD-7), the effect is small and the overall pattern of improvement remains robust.

Next, to determine whether the difference in distribution impacted our results, we re-estimated our models weighted according to the session distribution in the excluded group. The rate of depression symptom improvement was faster among the reweighted sample ($b_{reweighted}$=–1.55) compared with the unweighted sample (b=–1.49, Table 2), with similar results for anxiety improvement ($b_{reweighted}$=–1.34 vs b=–1.30). These results complement the results that included the number of sessions, suggesting that improvement may be faster when participants attend fewer sessions.

Finally, to ensure that our results were not upwardly biased by long intervals between appointments and assessments, we ran a sensitivity analysis where we restricted the interval to be closely tied in time: we excluded any assessments beyond 1 SD or 2 SD of the mean time between appointments and assessments and re-estimated the clinical improvement models. The rate of depression symptom improvement was faster (b_{1SD}=–1.54 and b_{2SD}=–1.56) when the interval was restricted compared with the rate of improvement without any restriction on assessment timing (b=–1.49, Table 2). A similar pattern was obtained for anxiety symptom improvement. This suggests that the true clinical outcomes of the program may be faster than what we observe when participants are not required to take assessments tied in time to their appointments.

Clinical Outcomes by Subgroup

Overview

To ensure representativeness, we compared our sample to the treatable population (ie, enrolled in benefits). Our sample differed from the population in race (χ^2_5=93.6, P<.001) and gender (χ^2_2=173.6, P<.001), but the effect sizes were negligible to small (Cohen w=0.08; w=0.13), reflecting more White and Hispanic patients and more women in the final sample than would be expected from population proportions. However, due to the extent of missingness of demographic data, race and gender subgroup analyses should be considered exploratory.

Race

For depression, there was a main effect of race, such that racial minority participants (except Asian) started therapy with higher baseline acuity than White participants (Table 3). Black and Hispanic participants improved faster than White participants (P<.011), and Asian participants of 2 or more races improved slower (P=.038). Yet, at 1-week post treatment, the overall effect size was large (raw prepost: d=1.55; TAU-adjusted d=0.91) and all participants' improvement (d values>0.69) was comparable or better than meta-analytic benchmarks (P<.3; Figure 4A). For anxiety, only Black participants started therapy with higher baseline acuity than White participants (Table 3). Asian participants improved faster than White participants (P=.008), and participants of 2 or more races improved more slowly (P=.034). Critically, at 1-week posttreatment, the overall effect size was large (raw prepost: d=1.72; TAU-adjusted d=1.13) and all participants' improvement (d values>0.87) was comparable or better to meta-analytic benchmarks (P values<.01; Figure 4B).

XSL•FO
RenderX

Table 3. Regression estimates of improvement in depression and anxiety symptoms by race (race and ethnicity data were provided by 30.7% of participants; Table 1).

	Depression (PHQ-9[a] score)	Anxiety (GAD-7[b] score)
Intercept		
Estimates	15.12	14.27
95% CI	15.01 to 15.23	14.17 to 14.37
P value	<.001	<.001
Race		
American Indian or Alaska Native		
Estimates	1.06	0.57
95% CI	0.01 to 2.11	–0.41 to 1.55
P value	.048	.256
Asian		
Estimates	–0.10	0.00
95% CI	–0.45 to 0.25	–0.32 to 0.32
P value	.576	.980
Black or African American		
Estimates	0.44	0.23
95% CI	0.20 to 0.68	0.01 to 0.45
P value	<.001	.044
Hispanic or Latino		
Estimates	0.37	0.19
95% CI	0.09 to 0.65	–0.07 to 0.45
P value	.009	.154
2 or more races/ethnicities		
Estimates	0.65	0.43
95% CI	0.12 to 1.17	–0.06 to 0.92
P value	.015	.083
Time in treatment (log-days)		
Estimates	–1.37	–1.18
95% CI	–1.39 to –1.34	–1.21 to –1.16
P value	<.001	<.001
Race (American Indian or Alaska Native) × Time in treatment		
Estimates	–0.17	0.07
95% CI	–0.41 to 0.07	–0.16 to 0.29
P value	.165	.550
Race (Asian) × Time in treatment		
Estimates	–0.04	–0.10
95% CI	–0.12 to 0.04	–0.18 to –0.03
P value	.376	.008
Race (Black or African American) × Time in treatment		
Estimates	–0.08	–0.04
95% CI	–0.14 to –0.02	–0.10 to –0.01
P value	.007	.110

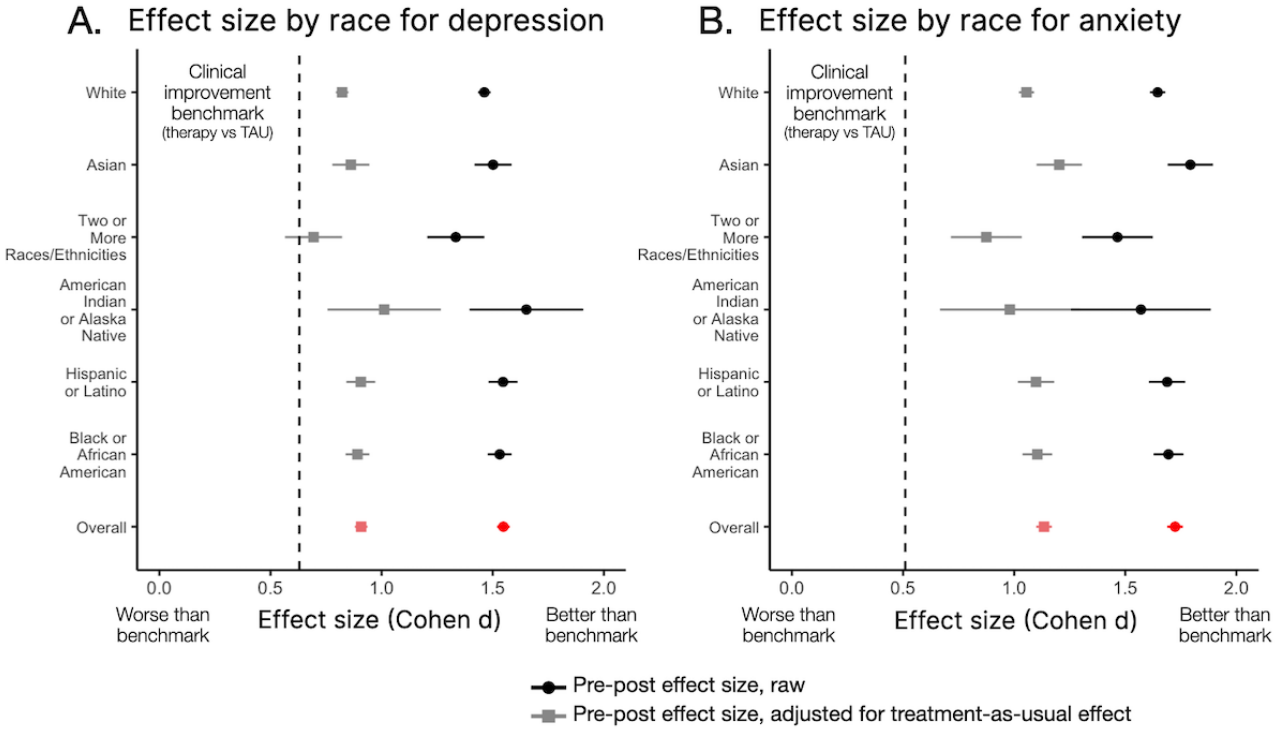
XSL•FO
RenderX

	Depression (PHQ-9[a] score)	Anxiety (GAD-7[b] score)
Race (Hispanic or Latino) × Time in treatment		
Estimates	–0.09	–0.04
95% CI	–0.15 to –0.02	–0.10 to 0.03
P value	.011	.258
Race (2 or more races/ethnicities) × Time in treatment		
Estimates	0.13	0.13
95% CI	0.01 to 0.25	0.01 to 0.24
P value	.038	.034
Random effects		
σ^2	15.55	12.99
τ_{00}		
Member_id	12.80	9.48
Intraclass correlation	0.45	0.42
N		
Member_id	13,296	12,244
Observations	44,628	41,454
Marginal R^2 / Conditional R^2	0.206 / 0.565	0.196 / 0.535

[a]PHQ-9: Patient Health Questionnaire 9-item Scale.

[b]GAD-7: Generalized Anxiety Disorder 7-item Scale.

Figure 4. The effect size of clinical improvement for different racial and ethnic minority participants, both for the prepost treatment effect (black) and when corrected for improvements expected due to TAU (gray). The TAU-adjusted effect sizes are plotted relative to meta-analytic effect sizes for psychotherapy versus TAU. (A) Improvement for depression, benchmarked against effect size of 0.63 (Cohen d). (B) Improvement for anxiety, benchmarked against effect size (d) of 0.51. All groups improved at or beyond the benchmarks. Race and ethnicity data were provided by 30.7% of participants. TAU: treatment as usual.



XSL•FO
RenderX

Gender

Men, women, and nonbinary participants' anxiety scores were not significantly different at the start of treatment, although nonbinary participants started more depressed (Table 4). For both depression and anxiety, nonbinary individuals improved more slowly than men (reference group; Table 4; *P* values<.003), and for anxiety, women also improved more slowly than men (*P*=.003). Critically, at 1 week post treatment, men and women's TAU-adjusted depression improvement (*d* values>0.77) was comparable or better than meta-analytic benchmarks (*P* values<.001) but nonbinary individuals' was slightly less (*d* difference=–0.10), although not significantly different from the benchmark (*z*=–1.24, *P*=.22). For anxiety, men, women and nonbinary participants' TAU-adjusted improvement (*d* values>0.65) was comparable or better than meta-analytic benchmarks (*P*<.20).

XSL•FO
RenderX

Table 4. Improvement in depression and anxiety symptoms by gender (gender data were provided by 18.9% of participants; Table 1).

	Depression (PHQ-9[a] score)	Anxiety (GAD-7[b] score)
Intercept		
Estimates	15.28	14.50
95% CI	15.05 to 15.50	14.29 to 14.71
P value	<.001	<.001
Gender		
Nonbinary		
Estimates	0.93	–0.08
95% CI	0.21 to 1.66	–0.77 to 0.60
P value	.012	.815
Woman		
Estimates	0.06	–0.17
95% CI	–0.20 to 0.32	–0.42 to 0.07
P value	.639	.158
Time in treatment (log-days)		
Estimates	–1.31	–1.24
95% CI	–1.36 to –1.25	–1.29 to –1.19
P value	<.001	<.001
Gender (Nonbinary) × time in treatment		
Estimates	0.24	0.37
95% CI	0.08 to 0.40	0.21 to 0.53
P value	.003	<.001
Gender (woman) × time in treatment		
Estimates	–0.06	0.09
95% CI	–0.12 to 0.00	0.03 to 0.14
P value	.065	.003
Random effects		
σ^2	15.55	12.88
τ_{00}		
Member_id	13.41	9.97
Intraclass correlation	0.46	0.44
N		
Member_id	8229	7610
Observations	29,390	27,395
Marginal R^2 / Conditional R^2	0.190 / 0.565	0.183 / 0.540

[a]PHQ-9: Patient Health Questionnaire 9-item Scale.

[b]GAD-7: Generalized Anxiety Disorder 7-item Scale.

Age

There was no difference in baseline severity for depression between participants aged 30 to 50 years, 50 to 65 years, and 65 years and older (*P* values>.31), but those aged 18 to 30 years were significantly more severe (*P*=.012; Table 5). For both depression and anxiety, younger age groups (30 to 50 years and 18 to 30 years) improved more slowly than the oldest (reference) age group, older than 65 years (Table 5; *P* values<.043). But at 1-week posttreatment, all participants had (TAU-adjusted) improvement greater than meta-analytic benchmarks (*d* values>0.84, *P* values<.001). For anxiety, there was no difference in baseline severity among age groups, and all

XSL•FO
RenderX

outperformed meta-analytic benchmarks (*d* values [TAU-adjusted] > 1.1, *P* values<.001).

XSL•FO
RenderX

Table 5. Improvement in depression and anxiety symptoms by age group.

	Depression (PHQ-9[a] score)	Anxiety (GAD-7[b] score)
Intercept		
Estimates	14.96	14.07
95% CI	14.58 to 15.33	13.69 to 14.45
P value	<.001	<.001
Age (years)		
50 to 65		
Estimates	0.22	0.20
95% CI	–0.18 to 0.61	–0.20 to 0.60
P value	.283	.326
30 to 50		
Estimates	0.22	0.28
95% CI	–0.16 to 0.60	–0.10 to 0.57
P value	.266	.146
18 to 30		
Estimates	0.50	0.35
95% CI	0.11 to 0.88	–0.04 to 0.74
P value	.012	.079
Time in treatment (log-days)		
Estimates	–1.56	–1.37
95% CI	–1.65 to –1.48	–1.46 to –1.29
P value	<.001	<.001
Age (50 to 65 years) × time in treatment		
Estimates	0.03	0.05
95% CI	–0.06 to 0.12	–0.04 to 0.14
P value	.528	.311
Age (30 to 50 years) × time in treatment		
Estimates	0.10	0.09
95% CI	0.02 to 0.19	0.00 to 0.18
P value	.020	.043
Age (18 to 30 years) × time in treatment		
Estimates	0.19	0.17
95% CI	0.10 to 0.27	0.08 to 0.26
P value	<.001	<.001
Random effects		
σ^2	15.56	12.84
τ_{00}		
Member_id	12.46	9.15
Intraclass correlation	0.44	0.42
N		
Member_id	42,630	39,860
Observations	137,607	128,966

XSL•FO
RenderX

	Depression (PHQ-9[a] score)	Anxiety (GAD-7[b] score)
Marginal R^2 / Conditional R^2	0.227 / 0.571	0.221 / 0.545

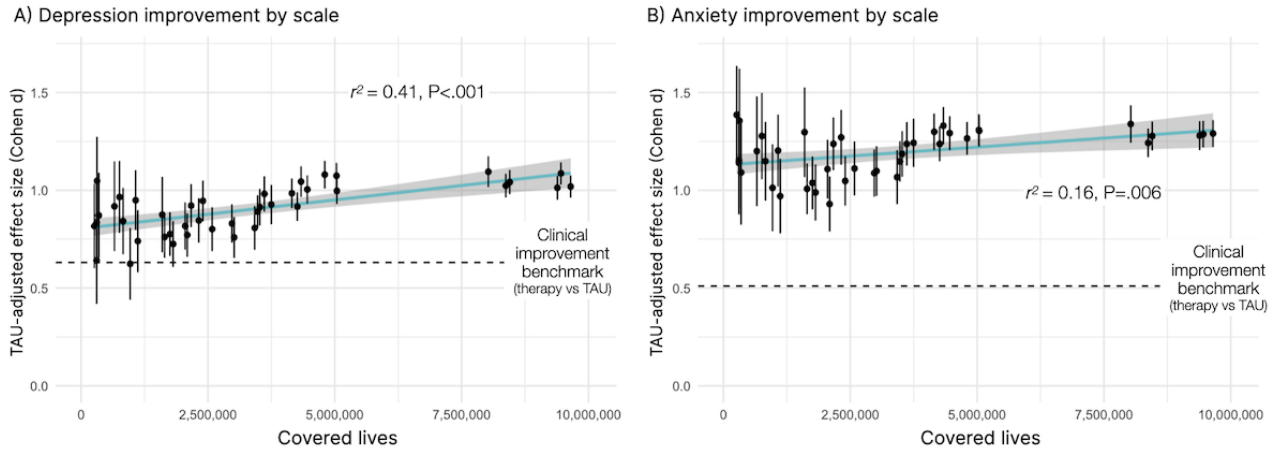
[a]PHQ-9: Patient Health Questionnaire 9-item Scale.

[b]GAD-7: Generalized Anxiety Disorder 7-item Scale.

Scaling's Impact on Outcomes

A total of 260,548 people were eligible for the benefit at the beginning of the study period, and 9,649,328 were eligible at the end, corresponding to a 37-fold increase in the number of people covered in 3 and a half years. Across the entire study period, 48% of those who enrolled and took an initial assessment started therapy. As the program scaled, this rate of conversion from assessment into care remained stable (b=0.00, P=.15), indicating treatment accessibility did not decline as a function of scale. Crucially, the effect size for symptom improvement for depression and anxiety increased relative to meta-analytic benchmarks as the program covered and provided treatment to more people ([per million]: depression: b=0.03, t=5.37, P<.001; anxiety: b=0.02, t=2.92, P=.006; Figure 5).

Figure 5. The effect size of clinical improvement (adjusted for improvement expected from TAU) for (A) depression and (B) anxiety as a function of program scale, with x-axis corresponding to how many lives were covered by the program. Improvements were benchmarked against a meta-analytic effect size (Cohen d) of 0.63 for depression and 0.51 for anxiety. The regression estimate is plotted in cyan, showing a significant increase in effect size as a function of scale. TAU: treatment as usual.



Discussion

Principal Findings

This large-scale cohort study—the largest of its kind to date—demonstrates how a comprehensive, employer-sponsored mental health program can facilitate timely access to care and leverage continuous data collection to deliver strong clinical outcomes for both depression and anxiety. Among the nearly 53,000 patients, logistic regression showed 92% of participants experienced reliable symptom improvement or recovery in depression or anxiety, and over 60% achieved remission, with PHQ-9 or GAD-7 scores dropping into the subclinical range, with similar results obtained when calculated directly by change in assessment scores. These results were robust to changes in assessment timing and number of sessions attended, supported through sensitivity analyses. This suggests low upward bias due to exclusion factors, although future work is needed to ensure more participants take reassessments at a regular cadence, which might be accomplished through more visible reminders or by providers during therapy sessions.

Benchmarking analysis found that outcomes met or significantly outperformed established meta-analytic standards for evidence-based treatment, both when considering the uncontrolled prepost effect sizes and when adjusting for expected improvement during TAU. These findings underscore the added value of an integrated, measurement-based approach to care.

Although a substantial portion of participants did not provide their race or gender, the availability of large-scale data enabled exploratory examination of clinical outcomes across diverse subgroups. Racial and ethnic minorities, who often confront systemic barriers to care, leading to lower quality or less evidence-based care [5,6] or provider biases [43], showed comparable rates of improvement to other groups, suggesting that the program provides appropriate support for patients from diverse backgrounds. Likewise, outcomes were similar among different age groups and similar between men and women, although nonbinary participants showed a somewhat smaller magnitude of improvement—an observation that may reflect stressors outside the individual's control, such as discrimination, stigma, and harassment, as proposed by minority stress theory [44]. Another possible factor could be the extent to which the treatment was tailored to meet the specific needs of nonbinary individuals. Since the current study lacks fine-grain gender data, future research is needed to understand these differences fully. Overall, across all subpopulations, the program outperformed or equaled gold-standard benchmarks [18,45]. Although these results are exploratory and cannot confirm equity, they suggest the program has potential to deliver high-quality care among

XSL•FO
RenderX

heterogeneous populations, as well as identifying gaps that future studies or program modifications must address (in particular, how to collect more comprehensive gender and race information without compelling patients to provide it).

Furthermore, these outcomes increased in magnitude as the program expanded, growing its coverage by 37 fold from the beginning of the study period to the end. Program features—including employer-sponsorship, a centralized care platform, care navigation, the ability for patients to rapidly find and choose their therapists, and consistent implementation of measurement-based care—likely contributed to sustaining access and improving clinical effectiveness. For example, therapy sessions were offered for free to patients as part of their benefit package, and the digital platform provided fast access, with an average wait time from assessment to appointment of about one day. Patients could select freely among available providers in the network, and whether they saw a full-time or contracted provider had no impact on their clinical improvement. The enhanced clinical outcomes may be attributed to the combined and interactive effects of multiple care components, rather than to any single intervention, which is typically assessed in benchmarking studies. This suggests that using a flexible range of evidence-based strategies could be crucial for the success of workplace mental health programs.

The current retrospective cohort study evaluated the therapeutic effectiveness of an employer-sponsored mental health program serving 53,000 patients treated by over 6800 providers. Overall, with broad demographic and clinical representation, the results demonstrate that a centralized, measurement-informed system of mental health care delivery can surpass standard benchmarks in improving depression and anxiety symptoms. Moving forward, further research should explore other components, such as how incentives to reward and retain the top-performing providers drive further clinical value.

Strengths and Limitations

This study has several strengths, such as a diverse and real-world sample, with 39% of participants (of the 16,250 participants who provided their race) being people of color, and a baseline comorbidity rate similar to the general population [39-41], although the results of this study, like other randomized controlled trials of psychotherapy, reflect different demographics than the United States in general (eg, skewing more White and more female). While we investigated broad race, age, and gender identity in our analysis, this approach does not capture the full spectrum of diversity characteristics that can influence therapy outcomes, and a substantial portion of participants opted not to share their race or gender information (69% missing race information; 81% missing gender information). Since missing data may introduce unknown and unmeasured bias into the results and constrain generalizability, the true magnitude of differences between different subgroups may vary, especially if the number of collected assessments differed substantially by demographic group.

Factors such as socioeconomic status, sexual orientation, cultural background, language proficiency, and disability status are integral to understanding individual experiences and treatment effectiveness [43]. As an observational cohort design that only examined outcomes before and after engaging in the program, this study cannot determine a causal link between improvements and engagement in the program. To better understand how measurement-based care contributes to program effectiveness, future interventions should drive higher follow-up assessment rates. Last, this study lacked a control group, although TAU effect sizes were subtracted from all uncontrolled prepost effect sizes to account for this limitation.

Conclusions

This retrospective cohort study demonstrates that a digital mental health benefit with a centralized system of care produces clinical outcomes in depression and anxiety equivalent to or significantly greater than what is typically observed through meta-analyses of psychotherapy. By using data to monitor, incentivize, and improve the quality of care. Furthermore, the clinical outcomes outperformed or equaled gold-standard benchmarks during a period of rapid growth in access to mental health care.

Data Availability

The datasets generated or analyzed during this study are not publicly available because this study comprises personal health information that is considered highly sensitive and confidential. These data are available upon request for researchers based on compliance with legal, regulatory, confidentiality, and patient privacy requirements (please contact the Yale Human Research Protection Program with requests [HRPP@Yale.edu]).

Authors' Contributions

EJW contributed to the conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, and visualization, as well as writing the original draft and reviewing and editing the manuscript. MH contributed to conceptualization, data curation, methodology, project administration, resources, supervision, and review and editing of the manuscript. MB contributed to conceptualization, resources, supervision, and review and editing of the manuscript. AMC contributed to conceptualization, supervision, and review and editing of the manuscript.

Conflicts of Interest

EJW, MH, MB, and AMC reported being employed by and holding equity in Spring Care Inc outside the submitted work. In addition, AMC reported being the lead inventor on 3 patent submissions relating to treatment for major depressive disorder (US

XSL•FO
RenderX

Patent and Trademark Office number Y0087.70116US00 and provisional application numbers 62/491 660 and 62/629 041) outside the submitted work. Finally, AMC reported holding equity in Carbon Health Technologies Inc, Wheel Health Inc, Parallel Technologies Inc, Healthie Inc, and UnitedHealthcare; receiving consulting fees from Fortress; and providing unpaid advisory services to health care technology startups outside the submitted work. No other disclosures were reported.

References

1. Depressive disorder (depression). World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2024-12-10]
2. Ferrari AJ, Charlson FJ, Norman RE, Flaxman AD, Patten SB, Vos T, et al. The epidemiological modelling of major depressive disorder: application for the global burden of disease study 2010. PLoS One 2013;8(7):e69637 [FREE Full text] [doi: 10.1371/journal.pone.0069637] [Medline: 23922765]
3. Mental health disorder statistics. Johns Hopkins Medicine. 2024. URL: https://www.hopkinsmedicine.org/health/wellness-and-prevention/mental-health-disorder-statistics [accessed 2024-12-10]
4. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO world mental health surveys. Psychol Med 2014;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
5. Primm AB, Vasquez MJT, Mays RA, Sammons-Posey D, McKnight-Eily LR, Presley-Cantrell LR, et al. The role of public health in addressing racial and ethnic disparities in mental health and mental illness. Prev Chronic Dis 2010;7(1):A20 [FREE Full text] [Medline: 20040235]
6. Wang PS, Berglund P, Kessler RC. Recent care of common mental disorders in the United States : prevalence and conformance with evidence-based recommendations. J Gen Intern Med 2000;15(5):284-292 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.9908044.x] [Medline: 10840263]
7. The 2024 NAMI workplace mental health poll. NAMI. 2024. URL: https://www.nami.org/support-education/publications-reports/survey-reports/the-2024-nami-workplace-mental-health-poll/ [accessed 2024-01-24]
8. Bondar J, Babich Morrow C, Gueorguieva R, Brown M, Hawrilenko M, Krystal JH, et al. Clinical and financial outcomes associated with a workplace mental health program before and during the COVID-19 pandemic. JAMA Netw Open 2022;5(6):e2216349 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.16349] [Medline: 35679044]
9. Ward EJ, Fragala MS, Birse CE, Hawrilenko M, Smolka C, Ambwani G, et al. Assessing the impact of a comprehensive mental health program on frontline health service workers. PLoS One 2023;18(11):e0294414 [FREE Full text] [doi: 10.1371/journal.pone.0294414] [Medline: 37988363]
10. Hawrilenko M, Smolka C, Ward E, Ambwani G, Brown M, Mohandas A, et al. Return on investment of enhanced behavioral health services. JAMA Netw Open 2025;8(2):e2457834 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.57834] [Medline: 39908020]
11. de Oliveira C, Saka M, Bone L, Jacobs R. The role of mental health on workplace productivity: a critical review of the literature. Appl Health Econ Health Policy 2023;21(2):167-193 [FREE Full text] [doi: 10.1007/s40258-022-00761-w] [Medline: 36376610]
12. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
13. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
14. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. Cogn Behav Pract 2015;22(1):49-59 [FREE Full text] [doi: 10.1016/j.cbpra.2014.01.010] [Medline: 27330267]
15. Khan BN, Liu RH, Chu C, Bolea-Alamañac B, Nguyen M, Thapar S, et al. Reach, uptake, and psychological outcomes of two publicly funded internet-based cognitive behavioural therapy programs in Ontario, Canada: an observational study. Int J Ment Health Syst 2024;18(1):34 [FREE Full text] [doi: 10.1186/s13033-024-00651-9] [Medline: 39511581]
16. Virtual solutions for depression and anxiety. Peterson Health Technology Institute. 2025. URL: https://phti.org/assessment/virtual-solutions-anxiety-depression/ [accessed 2025-07-29]
17. Cuijpers P, Harrer M, Miguel C, Ciharova M, Karyotaki E. Five decades of research on psychological treatments of depression: a historical and meta-analytic overview. Am Psychol 2025;80(3):297-310. [doi: 10.1037/amp0001250] [Medline: 37971844]
18. Papola D, Miguel C, Mazzaglia M, Franco P, Tedeschi F, Romero SA, et al. Psychotherapies for generalized anxiety disorder in adults: a systematic review and network meta-analysis of randomized clinical trials. JAMA Psychiatry 2024;81(3):250-259 [FREE Full text] [doi: 10.1001/jamapsychiatry.2023.3971] [Medline: 37851421]
19. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]

XSL•FO
RenderX

20. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
21. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
22. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
23. Lee KJ, Tilling KM, Cornish RP, Little RJA, Bell ML, Goetghebeur E, STRATOS initiative. Framework for the treatment and reporting of missing data in observational studies: the treatment and reporting of missing data in observational studies framework. J Clin Epidemiol 2021;134:79-88 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.01.008] [Medline: 33539930]
24. Flanagin A, Frey T, Christiansen SL, AMA Manual of Style Committee. Updated guidance on the reporting of race and ethnicity in medical and science journals. JAMA 2021;326(7):621-627. [doi: 10.1001/jama.2021.13304] [Medline: 34402850]
25. Lueger R, Barkham M. Using benchmarks and benchmarking to improve quality of practice and services. In: Developing and Delivering Practice-Based Evidence: A Guide for the Psychological Therapies. Hoboken New Jersey: Wiley Blackwell; 2010:223-256.
26. Minami T, Wampold BE, Serlin RC, Hamilton EG, Brown G, Kircher JC. Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: a preliminary study. J Consult Clin Psychol 2008;76(1):116-124. [doi: 10.1037/0022-006X.76.1.116] [Medline: 18229989]
27. Reese RJ, Duncan BL, Bohanske RT, Owen JJ, Minami T. Benchmarking outcomes in a public behavioral health setting: feedback as a quality improvement strategy. J Consult Clin Psychol 2014;82(4):731-742. [doi: 10.1037/a0036915] [Medline: 24841863]
28. Cuijpers P, Miguel C, Harrer M, Ciharova M, Karyotaki E. The overestimation of the effect sizes of psychotherapies for depression in waitlist controlled trials: a meta-analytic comparison with usual care controlled trials. Epidemiol Psychiatr Sci 2024;33:e56. [doi: 10.1017/S2045796024000611] [Medline: 39501660]
29. Cuijpers P, Quero S, Papola D, Cristea IA, Karyotaki E. Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: a meta-analysis. Psychol Med 2021;51(4):634-644. [doi: 10.1017/S0033291719003581] [Medline: 31843031]
30. Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, Greenhouse JB, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. application to the NIMH treatment of depression collaborative research program dataset. Arch Gen Psychiatry 1993;50(9):739-750. [doi: 10.1001/archpsyc.1993.01820210073009] [Medline: 8357299]
31. Diggle P. Analysis of Longitudinal Data. Oxford, United Kingdom: Oxford university press; 2002.
32. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. Wiley Online Library. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119513469 [accessed 2025-10-04]
33. Oehlert GW. A note on the Delta method. Am Stat 1992;46(1):27. [doi: 10.2307/2684406]
34. Kenward MG, Molenberghs G. Last observation carried forward: a crystal ball? J Biopharm Stat 2009;19(5):872-888. [doi: 10.1080/10543400903105406] [Medline: 20183449]
35. Jansen I, Beunckens C, Molenberghs G, Verbeke G, Mallinckrodt C. Analyzing incomplete discrete longitudinal clinical trial data. Statist. Sci 2006;21(1). [doi: 10.1214/088342305000000322]
36. Harrer M, Cuijpers P, Furukawa T, Ebert D. Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton,Florida: Chapman and Hall/CRC; 2021.
37. Hofmann SG, Kasch C, Reis A. Effect sizes of randomized-controlled studies of cognitive behavioral therapy for anxiety disorders over the past 30 years. Clin Psychol Rev 2025;117:102553 [FREE Full text] [doi: 10.1016/j.cpr.2025.102553] [Medline: 39970620]
38. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org/ [accessed 2025-10-04]
39. Brown TA, Campbell LA, Lehman CL, Grisham JR, Mancill RB. Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. J Abnorm Psychol 2001;110(4):585-599. [doi: 10.1037//0021-843x.110.4.585] [Medline: 11727948]
40. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the national comorbidity survey replication (NCS-R). JAMA 2003;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
41. Steffen A, Nübel J, Jacobi F, Bätzing J, Holstiege J. Mental and somatic comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data. BMC Psychiatry 2020;20(1):142 [FREE Full text] [doi: 10.1186/s12888-020-02546-8] [Medline: 32228541]
42. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol Methods 1997;2(1):64-78. [doi: 10.1037/1082-989X.2.1.64]
43. Huey S, Park A, Galán CA, Wang C. Culturally responsive cognitive behavioral therapy for ethnically diverse populations. Annu Rev Clin Psychol 2023;19:51-78 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080921-072750] [Medline: 36854287]

XSL•FO
RenderX

44. Frost DM, Meyer IH. Minority stress theory: application, critique, and continued relevance. Curr Opin Psychol 2023;51:101579 [FREE Full text] [doi: 10.1016/j.copsyc.2023.101579] [Medline: 37270877]
45. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008;76(6):909-922. [doi: 10.1037/a0013075] [Medline: 19045960]

Abbreviations

CBT: cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder 7-item Scale
PHQ-9: Patient Health Questionnaire 9-item Scale
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TAU: treatment as usual

Edited by E Mensah; submitted 24.02.25; peer-reviewed by C Chen; comments to author 07.04.25; revised version received 22.05.25; accepted 23.09.25; published 29.10.25.

Please cite as:
Ward EJ, Hawrilenko M, Brown M, Chekroud AM
Evidence-Based Mental Health at Scale: Benchmarking Retrospective Cohort Study of a Digital Employee Benefits Program for Depression and Anxiety
Online J Public Health Inform 2025;17:e72999
URL: https://ojphi.jmir.org/2025/1/e72999
doi:10.2196/72999
PMID:

©Emily J Ward, Matt Hawrilenko, Millard Brown, Adam M Chekroud. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 29.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Application of Machine Learning to Auto-Code Injury Data in the e-CHIRPP System: Development and Evaluation Study

Shamir N Mukhi[1], PhD; Steven R McFaull[2], MSc; Wendy Thompson[2], MSc; Tim Beattie[1], MSc

[1]Canadian Network for Public Health Intelligence, Public Health Agency of Canada, 9700 Jasper Ave, Edmonton, AB, Canada

[2]Canadian Hospitals Injury Reporting and Prevention Program, Centre for Surveillance and Applied Research, Public Health Agency of Canada, Ottawa, ON, Canada

Corresponding Author:
Shamir N Mukhi, PhD
Canadian Network for Public Health Intelligence, Public Health Agency of Canada, 9700 Jasper Ave, Edmonton, AB, Canada

Abstract

Background: The Canadian Hospitals Injury Reporting and Prevention Program (CHIRPP), a Public Health Agency of Canada program established in 1990, is an injury and poisoning sentinel surveillance system that collects and analyzes data on injuries to individuals who are seen at the emergency departments of numerous pediatric and general hospitals in Canada. Since its inception, the program has collected over 4 million records. The program's surveillance activities have contributed substantially to evidence-based decision-making to reduce injuries, support research, and establish preventive safeguards to protect the health and safety of Canadians. Patients presenting at participating hospitals are asked to complete a data collection form capturing the causes and circumstances contributing to the injury or poisoning event. Using this text, hospital and program staff have traditionally coded numerous surveillance variable codes manually for subsequent analysis within e-CHIRPP, the program's purpose-built analytical application on the Canadian Network for Public Health Intelligence public health informatics platform. Manual coding of this complex data is administratively burdensome and results in a significant time lag in the availability of important surveillance findings.

Objective: With the initial goal of achieving a preliminary stage of implementation, the objective was to establish the capability to achieve enhanced timeliness of surveillance findings within a process of adaptability and continuous improvement by applying machine learning to auto-code injury data based on patient narratives.

Methods: The research, development, and implementation of machine learning and auto-coding within the e-CHIRPP system were led by the Canadian Network for Public Health Intelligence team in collaboration with the CHIRPP program team. Data were extracted from e-CHIRPP and prepared for training, and candidate algorithms well suited for classification and supervised learning were initially assessed. Subsequently, 1 algorithm was chosen for further assessment based on initial accuracy, prediction confidence, and training time. The chosen algorithm was then further assessed in 2 stages, again using e-CHIRPP extracts: first, for a 2-year data set and then again for a 7-year data set. The sources of inaccuracies were investigated with a view to informing the refinement of the overall process and establishing ongoing adaptability and continuous improvement.

Results: Auto-coding of injury variables showed a high level of accuracy in most cases when compared to variables previously coded manually. Importantly, insights were also gained into the sources of observed inaccuracies and measures to foster ongoing refinement of the process.

Conclusions: The application of machine learning and auto-coding shows strong potential to benefit surveillance activities across various public health disciplines, yielding near real-time availability of intelligence, reduced administrative workload, continuous improvement, and adaptability to database growth and change.

(Online J Public Health Inform 2025;17:e69143) doi:10.2196/69143

KEYWORDS

injury; poisoning; surveillance; machine learning; auto-coding; public health; informatics

Introduction

Background

Injuries have a significant impact on people, institutions, and the economy in Canada. In 2021, a total of 24,034 Canadians died due to intentional and unintentional injuries, including poisonings. Unintentional injuries were the third leading cause of death in 2021, accounting for 6.2% of all deaths [1]. In addition to the lives lost, injuries have a negative impact on the economy due to lost productivity as well as increased health care costs. It has been estimated that injuries cost approximately CAD $29.4 billion for the Canadian economy in 2018, including CAD $20.4 billion in health care costs and approximately CAD

XSL•FO
RenderX

$9 billion in indirect costs [2]. To address the incidence of injuries and develop effective prevention strategies, timely and accurate information is needed to foster a clear understanding of their causes, the circumstances in which they occur, and the populations impacted, all of which can be captured through systematic collection of injury data [3].

The Canadian Hospitals Injury Reporting and Prevention Program

The Canadian Hospitals Injury Reporting and Prevention Program (CHIRPP), a Public Health Agency of Canada program established in 1990, is an injury and poisoning sentinel surveillance system that collects and analyzes data on injuries to individuals who are seen at the emergency departments (EDs) of numerous pediatric and general hospitals in Canada [4]. Since its inception, CHIRPP has collected over 4 million records, each capturing important details, such as how the injury happened, what the injured person was doing, circumstances and factors contributing to the incident, when and where it happened, as well as the age and sex of the injured person.

These surveillance activities have allowed CHIRPP to contribute substantially to evidence-based decision-making to reduce injuries, support research, and establish preventive safeguards. CHIRPP information has supported regulations on cribs, cradles, and bassinets [5]; magnetic toys [6]; vaping product labeling and packaging [7]; corded window coverings [8]; and safe storage of detergent packets [9]. The data have informed jurisdictional changes to all-terrain vehicle driving age [10] and changes to building codes for window heights in multistory complexes. In addition, the data have revealed injuries related to bath seats and informed safety measures related to children's play spaces and equipment, and sporting rules. Reports and publications have been produced on topics such as opioid-related ED visits before and during COVID-19 [11], self-harm during COVID-19 [12], e-cigarettes and vaping substances [13], as well as sports-related injuries and concussions [14].

Data collection is made possible through this longstanding collaborative partnership between participating hospitals and the CHIRPP program. In 2011, CHIRPP realized significant enhancements to the timeliness, flexibility, and adaptability of their surveillance tools through the implementation of the custom-developed e-CHIRPP system, a secure, web-based surveillance and data management technology on the Canadian Network for Public Health Intelligence (CNPHI) scientific public health informatics platform [15]. CNPHI and CHIRPP worked in collaboration to advance an innovative, dynamic code management system, which was developed and integrated within the initial release of e-CHIRPP to streamline how information was collected on the forms and aid in the adaptability of coding. This enabled the CHIRPP program team to maintain codes according to changing needs while aligning with associated coding rules and versioning, supporting improved coding and data integrity. While auto-coding of CHIRPP data was always a part of the proposed vision for CNPHI, e-CHIRPP established an important foundation for its future implementation.

Since then, there have been over a million injury cases captured in e-CHIRPP, yet despite the efficiencies gained, there are still manual tasks required of hospital staff and the CHIRPP program team. Currently, when a patient presents at a participating ED, the patient or caregiver is requested to complete the first part of a data collection form capturing the circumstances of the injury and how it occurred. Afterward, the attending physician or other hospital staff complete the second part of the data collection form, describing the nature of the injury, the treatment received, and any information on substance use, if applicable. Hospital staff code approximately half of the CHIRPP variables, creating draft cases that require further review by CHIRPP. After the draft cases are reviewed, coding is finalized by CHIRPP personnel. In general, coding that requires interpretation is completed by program staff, with the goal of describing each case as fully and consistently as possible, before being finalized.

Rationale

To accurately capture the wide array of causes and circumstances surrounding injury and poisoning events, CHIRPP uses a large number of codes that enable detailed analysis and intelligence generation via e-CHIRPP. For example, the variable *Intent* (capturing whether the injury was intentional or unintentional) has 8 codes to classify the information related to intent that corresponds to the *International Classification of Diseases, Tenth Revision*, coding, established by the World Health Organization. Additional examples include the variable *Location*, which uses 57 codes to accurately describe where the injury happened, and the variables *Direct cause* and *Factors 1 - 5*, each using the same 861 codes, thereby offering many layers to describe factors contributing to the injury. These and other variables are currently coded manually based on the patient narrative. Due to the complexity of CHIRPP data, the large number of codes, the large number of patients, and the ongoing possibility of adding new partner hospitals, there is an administrative burden resulting in a significant time lag between a patient's presentation at a hospital and the coding of the form used to complete a record on e-CHIRPP. Multiple studies in the past have also reported that manual coding of injury-related information and factors based on a free-text narrative is a time-consuming process [16]. The initial auto-coding work undertaken by CNPHI and CHIRPP commenced in 2019. At that time, the most recently available coded data were from 2017, a lag of 2 years. Table 1 presents the 18 CHIRPP variables and the number of codes used for each.

XSL•FO
RenderX

Table . Overview of CHIRPP[a] variables and related codes.

CHIRPP variable	Description	Number of codes
Indoor/outdoor[b]	Indicates whether the injury occurred indoors or outdoors.	3
Location	Describes where the injury generally occurred. It is used in conjunction with the area to further specify the room or named place where the injury/poisoning event occurred.	57
Area	Used to further specify within a general location where the injury/poisoning event occurred.	51
Context	Describes what the patient was doing or what activity they were engaged in when the injury/poisoning event occurred.	48
Vehicle type	The type of boat or motorized transport vehicle occupied by the patient in the injury/poisoning event.	13
Vehicle seating position	The vehicle seating position of the patient in the boat or motorized transport vehicle captured in vehicle type.	21
Direct cause[c]	The factor for which the most severe injury is attributed.	861
External cause	Describes both the intent and primary diagnosis (mechanism) of injury using the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*.	28
Sports/recreation	Any activity that requires physical exertion beyond baseline levels or an activity that is physical and uses a specific piece of equipment or is a physical game with a specific purpose.	120
Organized sport or activity	Indicates if the sport or physical activity that the patient was participating in at the time of the injury included coaches, instructors, or officials.	4
Safety equipment used	Identifies if safety equipment was used by the patient during the injury event.	3
Substance use	Indicates if illicit drugs, medications, alcohol, or other substances, either by the patient or another person, directly or indirectly caused the injury.	4
Intent	Describes the deliberate nature of the reported injury or poisoning by a human person. Injury or poisoning by the patient or another person is intentional when the injury or poisoning was the expected outcome.	8
Factors 1 to 5[d]	Inanimate objects, substances, and living things that directly or indirectly caused (contributed to) the injury/poisoning event.	861 each

[a]CHIRPP: Canadian Hospitals Injury Reporting and Prevention Program.

[b]Indoor/outdoor is binary; however, a third code captures instances where this is *unknown*.

[c]The direct cause code is chosen from the selection of factor codes. Coders first enter a factor code for the direct cause of the injury/poisoning corresponding to the most severe injury or poisoning, if known, and then may enter up to 5 additional factor codes, which also contributed to the injury/poisoning.

[d]Factors 1 to 5 provide additional layers to describe factors contributing to the injury.

Accordingly, there is a strong rationale to explore the applicability of machine learning and auto-coding of injury data with a view to greatly increasing the timeliness of intelligence arising from CHIRPP's surveillance. The approach can potentially assist in coding variables automatically, nearly in real time, as an innovative alternative to manual coding. In addition, the approach minimizes the need for human intervention, compared to manual coding, as the algorithm would be trained using predefined training data [17].

Objective

The objective of this initiative was to apply and assess the capability of machine learning to complete timely and accurate auto-coding of defined variable codes for injury and poisoning

XSL•FO
RenderX

events in e-CHIRPP based on information provided by patients or caregivers and to identify steps leading to preliminary implementation and ongoing refinement of the approach.

Methods

Overview

When a patient presents at a participating CHIRPP ED with an intentional or unintentional injury or poisoning, information is collected with the goal of fully describing the injury event and its context and location. For the purpose of this initiative, we used a machine learning algorithm to explore the feasibility of auto-coding 18 CHIRPP variables based on case narratives.

Ethical Considerations

Each hospital in the CHIRPP program has undertaken its own ethics review; no additional ethics review was required for this study. The data were drawn from an existing deidentified surveillance database and, therefore, anonymized. All methods were performed in accordance with the relevant guidelines and regulations.

Data Processing

Historical data from e-CHIRPP were extracted from the system. Two steps were performed, as described below.

Text tokenization was performed for the input variables. This is a process of dividing texts into words or smaller subtexts, determining the *vocabulary* of the data. This is like identifying various keywords of relevance that can be recognized by the algorithm and labeling the keywords with the appropriate codes.

This was followed by a feature selection process for the input variables. This is a process of removing irrelevant information that does not contribute to coding and prepares the dataset to be used for algorithm training.

Algorithm Selection

Overview

Numerous machine learning algorithms are available for an array of applications. The selection is not a question of which algorithm is best; it is a question of which is best suited for a particular type of application based on the dataset and the task at hand.

For the purpose of assessing algorithms to be used with CHIRPP data, numerous candidate algorithms were initially selected based on their suitability.

Classification

For application with CHIRPP data, machine learning is applied to predict which category the input data belongs to, whether the variables are binary (a or b), or multiclass (numerous).

Supervised Learning

This is a process of *teaching* an algorithm with labeled data (tokenized text) to prepare it for making predictions with real unforeseen data.

While a machine learning algorithm can be thought of as a computer-coded procedure for recognizing specific attributes or patterns within a dataset, a machine learning model comprises the output of an algorithm trained with labeled data and prepared for making predictions on unforeseen data [18].

The k-fold cross-validation method was used to assess the algorithms. This procedure involved randomizing the labeled data and then dividing it into *k* groups (folds) of approximately equal size. In each of a repeated series of trials, all but one of the folds are used as training data while one fold is used as a test data sample on which the algorithm makes predictions. Each fold is used once as the test data sample until the process is completed [19].

Evaluation of the algorithms was done using the following criteria: (1) accuracy (percentage of records that were classified correctly), (2) average confidence of all samples coded correctly (prediction confidence denoted by the model), (3) average confidence of all samples coded incorrectly (prediction confidence denoted by the model), and (4) training time (amount of time taken to train the model).

The purpose of this evaluation was to compare different machine learning algorithms and to select a single algorithm best suited for subsequent analysis. Multilayer perceptron (MLP) was selected as the preferred algorithm.

Evaluation of the MLP Algorithm

As a first step, data from e-CHIRPP for 2016 to 2017 were extracted from the system (316,803 cases). After removal of any draft cases, 80% of the remaining data were used as a training set to train the selected algorithm, and the remaining 20% of the data were used to test the algorithm. Using the injury event description or narrative, the auto-coding accuracy for each of the 18 CHIRPP injury variables was assessed.

To identify potential reasons for errors and misclassifications, a subset of samples that were misclassified was randomly selected and assessed manually.

The e-CHIRPP data contain codes that are used to indicate variables that are either unknown, irrelevant, or unable to be determined by the coder. Examples of these codes are *0IO: Missing* for the variable *indoor/outdoor* and *99A: unknown area* for the variable *area*. We refer to these codes as *null codes*. Simulations were repeated with null codes removed to assess the impact of potential *noise* removal and assess if the accuracy would improve.

This was followed by the extraction of 7 years of previously coded injury data from e-CHIRPP for 2011 to 2017 (976,249 cases). Similarly, after removal of any draft cases, 80% of the data were used as a training set to train the algorithm, and the remaining 20% were used to test the algorithm. The same analysis done on the initial 2 years of data was repeated on the 7-year dataset to determine whether an increase in the number of records had any impact on the accuracy of the auto-coding of the variables.

Given that the work of implementing machine learning and auto-coding was underway in 2019, the most recently coded data available at that time were from 2017 and earlier. It should be noted, however, that the arrival of the COVID-19 pandemic in 2020 impacted the pace of this initiative.

XSL•FO
RenderX

Results

Outcome Measures

The accuracy of auto-coding achieved by the algorithm was assessed by comparing the auto-coding results to the manually coded data, with % accuracy serving as an outcome measure based on the level of agreement between the two.

Primary Outcomes

Overall, we found that machine learning and auto-coding of injury variables showed a high level of accuracy in most cases when compared to variables previously coded manually. Importantly, we also gained insight into the potential sources of inaccuracies, as discussed in the secondary outcomes. Table 2 summarizes the primary outcomes achieved in all trials.

The % accuracy for auto-coding was highest for the variables *indoor/outdoor, substance use*, and *intent*, in general. This finding was consistent for both the 2- and 7-year datasets.

The accuracy was lowest for the variables *vehicle seating position* and *factors 1 to 5* for both datasets. We also found that the accuracy increased for all variables when null codes or missing values were removed, except for the variables *vehicle type* and *factors 4 and 5*.

In addition, when more data were used (7 years), the change in auto-coding performance was insignificant for most of the variables.

We also observed that the auto-coding accuracy progressively diminished in the results for *factor 1* through *factor 5*.

Table . Canadian Hospitals Injury Reporting and Prevention Program variables tested, number of codes for each variable, and % accuracy in auto-coding for the 2- and 7-y datasets.

Variables	Number of codes	2-y dataset including missing and null values (%)	2-y dataset missing/null values removed (%)	7-y dataset missing/null values removed (%)
Indoor/outdoor	3	89.96	95.32	95.36
Location	57	82.23	88.85	87.42
Area	51	87.19	93.33	91.25
Context	48	77.33	79.28	79.96
Vehicle type	13	98.84	84.91	85.18
Vehicle seating position	21	65.02	69.60	73.14
Direct cause	861	71.28	77.91	79.09
External cause	28	87.60	90.74	90.63
Sports/recreation	120	93.36	93.45	94.12
Organized sport or activity	4	85.30	94.60	92.54
Safety equipment used	3	79.84	92.40	92.40
Substance use	4	98.67	98.77	98.89
Intent	8	98.36	98.90	99.20
Factor 1	861	70.83	72.71	74.13
Factor 2	861	47.90	51.15	53.04
Factor 3	861	38.01	41.70	41.27
Factor 4	861	40.74	25.93	28.74
Factor 5	861	50.00	33.33	33.33

Secondary Outcomes

We manually assessed the cases that were incorrectly coded by the model to identify potential reasons for the misclassifications.

Some keywords extracted from the narratives did not have sufficient information at the time of coding. The narratives for some records were rich and consisted of many words or a few sentences, whereas some records had very limited narratives. This exercise helped to emphasize the importance of capturing a sufficient narrative from the patient or caregiver and hospital staff.

Some records had confusing narratives. For instance, for the variable "*indoor/outdoor*," there were some words in the narrative indicating that the injury occurred outdoors and some words indicating that the injury occurred indoors.

There were also inconsistencies in the historical, manually coded data used for training the algorithm. For instance, some coders coded injuries that occurred at the entrance of a building as indoors, whereas others coded them as occurring outdoors. Given the numerous codes related to some variables, this suggests that to some extent, the machine learning model revealed possible inconsistencies among human coders.

On investigating the observation that the auto-coding accuracy results progressively diminished from *factor 1* through *factor 5*, we observed an increasing scarcity in the use of these codes

XSL•FO
RenderX

as well as an increasing proportion of unpopulated data from *factor 1* through *factor 5* in the previously coded data.

Discussion

Strengths

Machine learning showed a promising capability for auto-coding complex injury and poisoning data derived from the free text narratives provided by patients or caregivers and hospital staff, with a high level of accuracy. The findings from this experience can serve to inform the successful implementation and ongoing refinement of the approach in this and other areas of public health surveillance and response.

The application of machine learning lends itself well to continuous improvement. With insights gained from the preliminary implementation of auto-coding and with supporting enhancements to the program's custom-built surveillance and analytical application on CNPHI, e-CHIRPP, the following features have been enabled: (1) randomized case selection for manual quality assurance verification of coded cases, with the rate of case review to be determined by the CHIRPP program; (2) flagging of cases coded with high confidence to recover examples of strong narratives to share with participating CHIRPP hospital staff for training and awareness purposes; and (3) flagging of cases of low confidence for manual review and to inform a cycle of algorithm retraining.

Limitations

Challenges Posed by the Large Number of Codes

We found that overall, the % accuracy for variables with a large number of codes, such as *factors 1 to 5* and *direct cause* (each using 861 codes), was lower than that of other variables. One possible explanation for this is that rarely used codes probably did not appear frequently enough in the extracted data. As a result, the algorithm did not have sufficient opportunity to learn the rarely used coding from the records [20].

Another possible explanation for low % accuracy is that due to the large number of codes, there is a higher likelihood of coding errors or inconsistencies among human coders. This would also impact the data used for algorithm training and testing. On the contrary, all the variables with less than 10 codes had an accuracy of greater than 90% for both the 2- and 7-year datasets without null codes, which supports the above reasoning.

Null or Missing Codes

The % accuracy increased for most of the variables when null and missing codes were removed. From a practical standpoint, it is undesirable for the algorithm to learn from data that are labeled as *unknown*. For example, the accuracy for auto-coding the variables *vehicle type* and *factors 4 and 5* decreased after null codes were removed because, in most cases, null codes appeared in the training data for these variables. This resulted in a significant decrease in the sample size available for algorithm training. The additional *factor* variables are available to describe cases involving more complex, multifactor circumstances contributing to an injury/poisoning event. While such cases are rare, it is important that the additional *factor* variables are available when needed, even if they are often not used. An increasing occurrence of null or missing codes in *factors 3* to 5 is an expected attribute of the data. For example, the CHIRPP team looked at a sample of coded cases (n=1,583,274) and found that, taken together, *factors 3* to 5 were used in less than 1% of the records. As a result, the remedy for null or missing codes aligns with that proposed for scarcely used codes, discussed below.

Scarcely Used Codes and Imbalanced Algorithm Training Data

When more data were used to make predictions (7 years), the change in performance for most variables was insignificant, which was initially an unexpected result. While it has been shown that the accuracy of auto-coding increases as the size of the training dataset increases, the incremental increase in accuracy eventually plateaus after a certain point, despite further increases in the size of the dataset [21]. However, rather than reaching a plateau, we suggest our results tend to highlight the degree of scarcity in the use of some codes because the use of additional data had little impact on accuracy, particularly for the variables *direct cause* and *factors 1 to 5*, which all use the same 861 codes. While additional volumes of training data are recommended for further rounds of training and assessment, future remedies to address scarcely used codes could also include the use of artificial training data to target the most scarcely used codes not encountered in training data.

One of the biggest challenges we faced was with imbalanced datasets where some codes appeared significantly more frequently than others. Some variables (*vehicle type*, *substance use*, and *intent*) have highly skewed datasets where >90% of all cases have *dominant* codes. The drawback of this is that if a certain variable is highly skewed, then a model could simply always predict the dominant codes, and it would still score a high accuracy, while remaining unsuccessful in learning the nondominant codes. We suggest that the best way to deal with this is to rebalance the dataset by removing some of the dominant codes to achieve a dataset with 50% dominant and 50% nondominant codes for training purposes. This may have the effect of lowering the overall accuracy slightly but can significantly increase the accuracy of auto-coding nondominant codes. Various approaches to rebalancing datasets continue to evolve in the area of machine learning, which can be incorporated as a part of the continuous improvement of the overall process [22].

Subjectivity in Recalling or Interpreting Details

Information provided in patient narratives may be impacted by subjectivity, varying levels of detail, or differing interpretations among coders. This may partially explain the low initial results for accuracy in auto-coding the variable *context*, for example. There is a potential that inconsistencies are introduced at these stages that have a subsequent impact on algorithm training and the resulting accuracy of auto-coding.

Data Years Used and Pandemic-Related Impacts

Although our work commenced in 2019, using the most recently available coded data from 2017 and earlier, the project was impacted by the arrival of the COVID-19 pandemic. While the objective of reaching a preliminary stage of implementation of

XSL•FO
RenderX

auto-coding injury data was achieved, further stages of refinement and validation have been informed by this initial experience.

A Human Resource Is Still Required

Auto-coding using machine learning is not perfect and does not fully eliminate the need for a human resource. In the future, a blended approach could involve human investigation of auto-coding results of low confidence or those flagged for attention by the model and verification of samples of auto-coded cases. Manual investigation and coding could correct such instances, supporting ongoing refinement of the process.

Conclusions

This innovation has demonstrated the successful preliminary implementation of a solution to address the administrative burden and time lag associated with the manual coding of large quantities of complex surveillance data. Other public health disciplines could also benefit from the application of machine learning to realize near real-time availability of surveillance intelligence, fostering improved timeliness of actions and decisions to protect public health. Importantly, the process readily lends itself to ongoing adaptability and continuous improvement, offering capabilities to accommodate new or changing information fields and address sources of inaccuracy. While the objective of reaching a preliminary stage of implementation was achieved, challenges related to the collection and auto-coding of complex injury data have been characterized to inform future steps for the partners.

A strategy of ongoing verification and process review is required to ensure optimal accuracy and continuous improvement of the approach. Quality assurance and verification activities need to be customized on a program-specific basis. A human resource is still required to oversee and maintain the process. A surveillance and analytical application suitable for a program's needs, as in the case of CHIRPP's e-CHIRPP system on CNPHI, provides an important, complementary foundation that is instrumental in supporting continuous improvement.

Insights gained from this preliminary stage of implementation will position the partners to embark on the next steps, leading to an envisioned state where time-consuming human coding can be replaced with auto-coding supported by a rigorous strategy of quality assurance, validation, and refinement. This could include the targeted review of records below set confidence levels and randomized review of a set percentage of coded cases on an ongoing basis.

Despite some of the complexities encountered, there are many plausible benefits of auto-coding CHIRPP variables using machine learning. These benefits extend to other public health surveillance disciplines as well and may include the following:

- *Timely availability of intelligence to identify and respond to public health risks and emerging issues:* Auto-coding using machine learning could potentially provide comprehensive information in near real-time, largely eliminating lag and enabling more timely responses to protect public health.
- *Accuracy of the coding:* Algorithms can potentially improve the accuracy of the coded data by decreasing human error and eliminating inconsistencies among coders, resulting in more precise surveillance intelligence.
- *Reduction in workload for coders:* Manual coding based on a free text narrative, especially for large quantities of complex data, is very resource intensive, leading to a heavy workload for coders. Auto-coding could reduce this workload, save time, and potentially increase their productivity.
- *Database growth and addition of variables:* Manual coding becomes more challenging and complex as a system grows and new variables or codes are added. Machine learning algorithms could adapt to such growth with minimal challenges.
- *Resource allocation and program planning:* Intelligence from a completely coded database could support program planning, resource allocation, and research while informing preventive strategies to protect public health.

Acknowledgments

The authors would like to acknowledge Tim Tran for assisting with the research and development of the models, as well as Karna Dhiravani, Reem Maarouf, Lionel Pineda, Lin Yan, and the rest of the CNPHI team. The authors would also like to acknowledge the contributions of James Cheesman, Calvin Gale, Brian Wray, the Canadian Hospitals Injury Reporting and Prevention Program hospitals, and the Canadian Hospitals Injury Reporting and Prevention Program surveillance team.

Authors' Contributions

SM: conceptualization, methodology, project administration, writing – original draft, writing – review & editing
WT: conceptualization, methodology, writing – review & editing
SRM: conceptualization, methodology, validation, writing – review & editing
TB: writing – original draft, writing – review & editing

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Table 13-10-0156-01 deaths, by cause, chapter XX: external causes of morbidity and mortality. Statistics Canada. 2023. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310015601 [accessed 2025-11-26]
2. Parachute. The cost of injury in Canada. 2021. URL: https://parachute.ca/en/professional-resource/cost-of-injury-in-canada/ [accessed 2025-11-26]
3. Injury surveillance guidelines (nowho/NMH/VIP/0102). : World Health Organization; 2001 URL: https://iris.who.int/server/api/core/bitstreams/55f4d389-71c6-4677-81e6-07847f7507d8/content [accessed 2025-12-05]
4. Crain J, McFaull S, Thompson W, et al. Status report - The Canadian Hospitals Injury Reporting and Prevention Program: a dynamic and innovative injury surveillance system. Health Promot Chronic Dis Prev Can 2016 Jun;36(6):112-117. [doi: 10.24095/hpcdp.36.6.02] [Medline: 27284703]
5. Cribs, cradles and bassinets regulations: SOR/2016-152. Canada Gazette. 2016. URL: https://gazette.gc.ca/rp-pr/p2/2016/2016-06-29/html/sor-dors152-eng.html [accessed 2025-11-26]
6. Regulations amending the toy regulations (magnetic toys): SOR/2018-138. Canada Gazette. 2018. URL: https://gazette.gc.ca/rp-pr/p2/2018/2018-07-11/html/sor-dors138-eng.html [accessed 2025-11-26]
7. Vaping products labelling and packaging regulations: SOR/2019-353. Canada Gazette. 2019. URL: https://gazette.gc.ca/rp-pr/p2/2019/2019-12-25/html/sor-dors353-eng.html [accessed 2025-11-26]
8. Notice to stakeholders – new requirements for corded window coverings. Government of Canada. 2021. URL: https://www.canada.ca/en/health-canada/services/consumer-product-safety/legislation-guidelines/guidelines-policies/notice-stakeholders-corded-window-coverings.html [accessed 2025-11-26]
9. Zutrauen S, Cheesman J, McFaull SR. Pediatric injuries and poisonings associated with detergent packets: results from the Canadian Hospitals Injury Reporting and Prevention Program (CHIRPP), 2011-2023. Inj Epidemiol 2024 Jul 11;11(1):31. [doi: 10.1186/s40621-024-00513-5] [Medline: 38992784]
10. McLean L, Russell K, McFaull S, Warda L, Tenenbein M, McGavock J. Age and the risk of all-terrain vehicle-related injuries in children and adolescents: a cross sectional study. BMC Pediatr 2017 Mar 17;17(1):81. [doi: 10.1186/s12887-017-0807-y] [Medline: 28302103]
11. Yao X, McFaull SR, Champagne AS, Thompson W. An interrupted time series analysis of trends in opioid-related emergency department visits from pre-COVID-19 pandemic to pandemic, from the Canadian Hospitals Injury Reporting and Prevention Program. BMC Public Health 2023 Aug 4;23(1):1483. [doi: 10.1186/s12889-023-16414-z] [Medline: 37537534]
12. Saxena S, Liu L, Pollock N, McFaull SR. Self-harm emergency department visits in Canada during the COVID-19 pandemic: evidence from a sentinel surveillance system. Inj Epidemiol 2023 Jan 5;10(1):1. [doi: 10.1186/s40621-022-00411-8] [Medline: 36604736]
13. McFaull SR, Do MT, Champagne A, Bang F. Injuries and poisonings associated with e-cigarettes and vaping substances, electronic Canadian Hospitals Injury Reporting and Prevention Program, 2011-2019. Health Promot Chronic Dis Prev Can 2020 Aug;40(7-8):250-254. [doi: 10.24095/hpcdp.40.7/8.05] [Medline: 32667882]
14. Fridman L, Fraser-Thomas JL, McFaull SR, Macpherson AK. Epidemiology of sports-related injuries in children and youth presenting to Canadian emergency departments from 2007-2010. BMC Sports Sci Med Rehabil 2013 Dec 23;5(1):30. [doi: 10.1186/2052-1847-5-30] [Medline: 24364875]
15. Mukhi S, Aramini J, Kabani A. Contributing to communicable diseases intelligence management in Canada: CACMID meeting, March 2007, Halifax, Nova Scotia. Can J Infect Dis Med Microbiol 2007 Nov;18(6):353-356. [doi: 10.1155/2007/386481] [Medline: 18978985]
16. Vallmuur K, Marucci-Wellman HR, Taylor JA, Lehto M, Corns HL, Smith GS. Harnessing information from injury narratives in the “big data” era: understanding and applying machine learning for injury surveillance. Inj Prev 2016 Apr;22 Suppl 1(Suppl 1):i34-i42. [doi: 10.1136/injuryprev-2015-041813] [Medline: 26728004]
17. Lehto M, Marucci-Wellman H, Corns H. Bayesian methods: a useful tool for classifying injury narratives into cause groups. Inj Prev 2009 Aug;15(4):259-265. [doi: 10.1136/ip.2008.021337] [Medline: 19652000]
18. Brownlee J. Machine learning mastery difference between algorithm and model in machine learning. 2020. URL: https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/ [accessed 2025-11-26]
19. Brownlee J. A gentle introduction to k-fold cross validation. Machine Learning Mastery. 2018. URL: https://machinelearningmastery.com/k-fold-cross-validation/ [accessed 2025-11-26]
20. Vallmuur K. Machine learning approaches to analysing textual injury surveillance data: a systematic review. Accid Anal Prev 2015 Jun;79:41-49. [doi: 10.1016/j.aap.2015.03.018]
21. Zargaran E, Spence R, Adolph L, et al. Association between real-time electronic injury surveillance applications and clinical documentation and data acquisition in a South African trauma center. JAMA Surg 2018 May 16;153(5):e180087. [doi: 10.1001/jamasurg.2018.0087] [Medline: 29541765]
22. Susan S, Kumar A. The balancing trick: optimized sampling of imbalanced datasets—a brief survey of the recent state of the art. Eng Rep 2021 Apr;3(4):e12298 [FREE Full text]

Abbreviations

CHIRPP: Canadian Hospitals Injury Reporting and Prevention Program

XSL•FO
RenderX

CNPHI: Canadian Network for Public Health Intelligence
e-CHIRPP: the purpose-built analytical public health informatics application used by the CHIRPP team
ED: emergency department

Edited by E Mensah; submitted 22.11.24; peer-reviewed by C Asuzu, J Taylor, ML Zwald; revised version received 06.08.25; accepted 11.11.25; published 17.12.25.

Please cite as:
Mukhi SN, McFaull SR, Thompson W, Beattie T
Application of Machine Learning to Auto-Code Injury Data in the e-CHIRPP System: Development and Evaluation Study
Online J Public Health Inform 2025;17:e69143
URL: https://ojphi.jmir.org/2025/1/e69143
doi:10.2196/69143

© Shamir N Mukhi, Steven R McFaull, Wendy Thompson, Tim Beattie. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 17.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Forecasting Neonatal Mortality in Ethiopia to Assess Progress Toward National and International Reduction Targets Using Classical Techniques and Deep Learning: Time-Series Forecasting Study

Shimels Derso Kebede, MPH

Department of Health Informatics, School of Public Health, College of Medicine and Health Science, Wollo University, South Wollo Zone, Amhara Region, Dessie, Ethiopia

Corresponding Author:
Shimels Derso Kebede, MPH
Department of Health Informatics, School of Public Health, College of Medicine and Health Science, Wollo University, South Wollo Zone, Amhara Region, Dessie, Ethiopia

Abstract

Background: Neonatal disease and its outcomes are important indicators for a responsive health care system and encompass the effects of socioeconomic and environmental factors on new-borns and mothers. Ethiopia is working to achieve the Sustainable Development Goal target for the reduction of 12 or less per 1000 birth by 2030 and 21 per 1000 livebirths by 2025 as part of the second Ethiopian Health Sector Transformation Plan.

Objective: This study aimed to compare the performance of classical time-series models with that of deep learning models and to forecast the neonatal mortality rate in Ethiopia to verify whether Ethiopia will achieve national and international targets.

Methods: Data were extracted from the official World Bank database. Classical time-series models, such as autoregressive integrated moving average (ARIMA) and double exponential smoothing, and neural network-based models, such as multilayer perceptron, convolutional neural network, and long short-term memory, have been applied to forecast neonatal mortality rates from 2021 to 2030 in Ethiopia. During model building, the first 21 years of data (from 1990 to 2010) were used for training, and the remaining 10 years of data were used to test model performance. Model performance was evaluated using R^2, mean absolute percentage error (MAPE), and root mean squared error (RMSE). Finally, the best model was used to forecast the neonatal mortality rate over the next 10 years from 2021 to 2030, with a 95% prediction interval (PI).

Results: The results showed that the double exponential smoothing model was the best, with a maximum R^2 of 99.94% and minimum MAPE and RMSE of 0.002 and 0.0748, respectively. The worst performing among the 5 models was the CNN, with an R^2 of 93.71% and a maximum RMSE of 0.79. Neonatal mortality in Ethiopia is forecasted to be 23.20 (PI 22.20-24.40) per 1000 live births in 2025 and 19.80 (PI 17.10-22.80) per 1000 live births in 2030.

Conclusions: This study revealed that national and international targets for neonatal mortality cannot be realized if the current trend continues. This highlights the need for urgent interventions to strengthen the health system to fasten the decline rate of neonatal mortality and collaborative effort with concerned stakeholders for improved and responsive neonatal and child health services in order to achieve these targets.

(Online J Public Health Inform 2025;17:e66798) doi:10.2196/66798

KEYWORDS

neonatal mortality; time series; forecasting; deep learning; machine learning; health sector transformation plan; deep learning model; Ethiopia

Introduction

The neonatal mortality rate refers to the number of deaths during the first 28 days of life per 1000 live births, and it remains a significant global public health problem globally [1]. Neonatal diseases and their outcomes are important indicators of responsive health care systems and encompass the effects of socioeconomic and environmental factors on newborns and mothers. Although prevention and treatment are simple, relatively inexpensive, and efficient, the contribution of neonatal mortality to overall mortality and morbidity is overwhelming and unequaled by any other condition, as no age group has received less attention than neonates. Notwithstanding efforts to achieve the United Nations' Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs) over the

XSL•FO
RenderX

past 2 decades, children are still most likely to die during the neonatal period [2].

In 2019, 70% of deaths among children and youth aged 25 years and younger occurred in those aged 5 years and younger, with neonatal deaths accounting for a larger share of under-5 mortality. The contribution of neonatal mortality to all under-5 deaths increased from 40% in 1990 to 47% in 2019 [3]. Globally, 53 countries will not achieve the SDG target for under-5 mortality by 2030 if the trend evident from 2010 to 2019 continues. To meet these targets, resources and policies must be mobilized to accelerate progress rather than sustain the current declining trend.

Many low- and middle-income countries have reduced the number of children who die before the age of 5 years as a result of interventions undertaken since 1990 to meet MDG 4, which aimed at a two-thirds reduction in the mortality risks of children ages 5 years and younger by 2015. Despite the success achieved by MDGs in low- and middle-income countries, neonatal mortality is still unacceptably high, with the largest burden evident in sub-Saharan African countries [4]. According to a United Nations International Children's Emergency Fund (UNICEF) report, sub-Saharan African neonates were at tenfold greater risk of death than neonates born in developed countries. The report also showed that most neonatal deaths occurred in East Africa, more than 50% of them in Ethiopia [5].

In Ethiopia, a lower decline of 41% was observed in neonatal mortality compared to the 60% reduction in under-5 mortality and 50% reduction in infant mortality between 2000 and 2016 [6]. According to the SDG index, Ethiopia ranks 136 out of 166 countries, with an overall performance score of 55.2, which is above the regional average score of 53.1 [7]. This makes achievement of the SDG target to reduce the neonatal mortality rate to 12 per 1000 births by 2030, and the second Ethiopian Health Sector Transformation Plan (HSTP-II) target to reduce the neonatal mortality rate to 21 per 1000 live births by 2025 [8], questionable without special effort being made.

In this study, deep-learning models were applied in addition to classical time-series approaches to forecasting. Deep-learning models involve features such as the ability to handle temporal dependencies, distribution-free learning, and flexibility when modeling nonlinear features, which make them suitable for time-series data modeling [9]. This study aimed to compare the performance of commonly known classical time-series models with deep-learning models when forecasting the annual neonatal mortality rate in Ethiopia to verify whether Ethiopia achieves national and international targets. To the best of our knowledge, this paper presents a pioneering exploration of neonatal mortality forecasting using classical time-series and deep-learning techniques. Using historical data to predict future neonatal mortality is important for understanding trends and estimating the burden of neonatal mortality. The findings will help track Ethiopia's progress toward achieving the SDG by 2030 and the national HSTP-II target by 2025.

Methods

Data Source

Neonatal mortality data used in this study were obtained from the World Bank's Health Nutrition and Population Statistics database [10]. These data are compiled by the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME), which includes the UNICEF, the World Health Organization, the World Bank, and the United Nations Department of Economic and Social Affairs (Population Division). The UN IGME uses a broad strategy to arrive at annual estimates of child mortality by compiling and assessing the quality of all available nationally representative data relevant to the estimation of child mortality, including data from vital registration systems, population censuses, household surveys, and sample registration systems [3]. Similarly, the neonatal mortality estimates are derived using standardized statistical models that synthesize data from a variety of sources, including national censuses, household surveys (such as Demographic and Health Surveys and Multiple Indicator Cluster Surveys), vital registration systems, and other country-reported data. These estimates are produced following rigorous validation methods to ensure comparability across countries and over time. For this study, we extracted Ethiopia's national-level annual neonatal mortality rates for the period 1990 to 2020. These figures represent model-based estimates of the overall neonatal mortality rate in the country, expressed as the number of deaths within the first 28 days of life per 1000 live births per year. The data are publicly available through the UN IGME platform at [11].

Data Processing and Analysis

Autoregressive integrated moving average (ARIMA) and double exponential smoothing from classical time-series models and multilayer perceptron (MLP), deep-learning algorithms such as convolutional neural networks (CNN), and long short-term memory (LSTM) were selected to forecast the neonatal mortality rate from 2021 to 2030 in Ethiopia.

Classical Models

Classical time-series forecasting models, such as ARIMA and exponential smoothing, were applied using the well-known Python (Python Software Foundation) package, *statsmodels*.

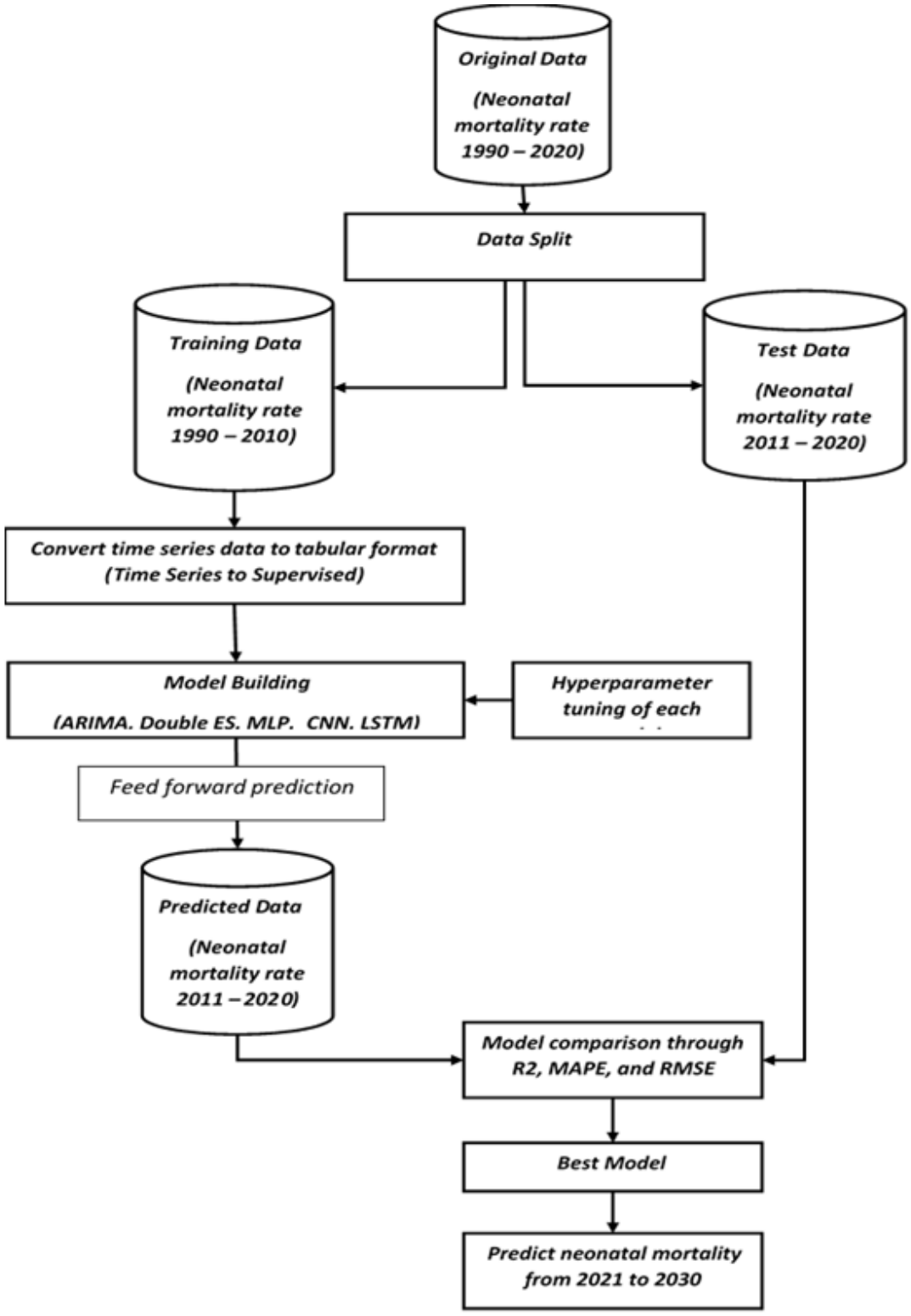
ARIMA is a class of statistical models for analyzing and forecasting time-series data that incorporates the concept of integration and is a generalization of the straightforward autoregressive moving average. The autoregression (AR) part of ARIMA shows that the time series is regressed on its own past observations, the moving average part indicates that the forecast error is a linear combination of past respective errors, and the integrated (I) part shows that the data values have been replaced with different values of d to remove trends in the dataset and obtain stationary data, which is the assumption of the ARIMA model [12]. ARIMA forecasting is capable of dealing with nonstationary time-series data because of its "integrate" component that involves differencing the time series to convert a nonstationary time series into a stationary time series, which means that statistical properties such as mean, variance, autocorrelation, etc, are all constant over time. The

XSL•FO
RenderX

general form of the ARIMA model is denoted as ARIMA (p, d, q) to represent the AR, I, and moving average components, respectively [13]. The difficulty of specifying p, d, and q (which is a significant limitation) can be overcome by experimenting with various combinations and assessing the model's performance [12]. Hence, instead of using subjective diagnostic visualizations, such as the autocorrelation function and partial autocorrelation function, to determine these parameters (p, d, q), grid searching of specified values was performed, and the best combination that minimized the forecasting error was used to determine the best ARIMA model. ARIMA, which is a time series with a repeating cycle, has the drawback of not supporting the seasonal data. ARIMA expects data that are not seasonal or have had the seasonal component eliminated, such as data that have been seasonally adjusted using techniques such as seasonal differencing [14]. Fitting seasonal ARIMA (SARIMA), a version of the ARIMA model that incorporates additional seasonal components, is an alternative approach. However, Figure 1 shows that the data shows no repeating patterns.

XSL•FO
RenderX

Figure 1. Workflow of data preparation and analysis plan to forecast neonatal mortality in Ethiopia. ARIMA: autoregressive integrated moving average; ES: Exponential smoothing; MLP: multilayer perceptron; CNN: convolutional neural network; LSTM: long short-term memory; MAPE: mean absolute percentage error; RMSE: root mean squared error.



The mathematical form of an ARIMA (p, d, q) model can be written as follows [15]:

$$y_t' = c + \varphi_1 y_{t-1}' + \varphi_2 y_{t-2}' + \cdots + \varphi_p y_{t-p}' + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where:

y_t' is the differenced series after applying the integration part d times (ie, $y_t' = y_t - y_{t-1}$, if $d=1$),

c is a constant,

$\varphi_1, \varphi_2, \ldots, \varphi_p$ are the coefficients for the autoregressive terms,

$\theta_1, \theta_2, \ldots, \theta_q$ are the coefficients for the moving average terms,

ε_t is the error term (white noise) at time t,

ε_{t-1}, ε_{t-2}, … are past forecast errors.

XSL•FO
RenderX

Another classical time-series model used in this study is exponential smoothing. Time-series methods, such as the ARIMA family of methods, develop a model in which the prediction is a weighted linear sum of recent past observations. Exponential smoothing forecasting methods are similar in that a prediction is a weighted sum of past observations, but the model explicitly uses an exponentially decreasing weight for past observations; the more recent the observation, the higher is the associated weight [14]. There are 3 main types of exponential smoothing time-series forecasting methods: simple exponential smoothing, a simple method that assumes no systematic structure (ie, without a trend or seasonality); double exponential smoothing, an extension that explicitly handles trends; and triple exponential smoothing, which adds support to seasonality. Because our dataset had only a trend component, a double exponential smoothing model was fitted.

Double exponential smoothing requires parameters called alpha (α), also called the smoothing factor or smoothing coefficient, which controls the rate at which the influence of observations at previous time steps decays exponentially. α is often set between 0 and 1. A value close to 1 indicates fast learning (ie, only the most recent values influence the forecasts), whereas a value close to 0 indicates slow learning (ie, past observations have a large influence on forecasts). In addition to the α parameter, an additional smoothing factor, called beta (b or β), is added to control the decay of the influence of the change in trend. The method supports trends that change in different ways, additive and multiplicative, depending on whether the trend is linear or exponential, respectively. The damping coefficient Phi (p or ϕ) is another hyperparameter used to control the damping rate. Dampening means reducing the size of the trend over future time steps to a straight line (no trend), and it can be additive or multiplicative for a linear or exponential dampening effect. In general, double exponential smoothing has five hyperparameters: (1) alpha (α): smoothing factor for the level, (2) beta (β): smoothing factor for the trend, (3) trend type: additive or multiplicative, (4) dampen type: additive or multiplicative, and (5) phi (ϕ): damping coefficient.

The mathematical form of double exponential smoothing model can be written as follows [15]:

$$y\hat{}t+h\vee t=lt+(\varphi+\varphi 2+\ldots+\varphi h)bt$$

$$lt=\alpha yt+(1-\alpha)(lt-1+\varphi bt-1)$$

$$bt=\beta(lt-lt-1)+(1-\beta)\varphi bt-1$$

Where: ℓt denotes an estimate of the level of the series at time t,

bt denotes an estimate of the trend (slope) of the series at time

t, h denotes how many periods into the future we're forecasting,

α is the smoothing parameter for the level, $0\leq\alpha\leq 1$,

β is the smoothing parameter for the trend, $0\leq\beta\leq 1$, and

φ is the damping parameter $0<\varphi<1$

Among them, the optimal values of alpha, beta, and phi can be automatically tuned by the model, and the optimal values of the trend- and dampen-type dampens are optimized by grid search hyperparameter tuning, as they could not be optimized automatically by the model.

Deep Learning Algorithms for Time-Series Forecasting

Owing to their familiarity and versatility, linear techniques such as ARIMA have historically dominated time-series forecasting. However, these traditional approaches have certain drawbacks, such as the exclusion of complex joint distributions by assuming a linear relationship. However, machine-learning techniques can be useful for solving complex time-series forecasting issues that involve multiple input variables, sophisticated nonlinear relationships, and missing data.

In this study, the most basic type of feed-forward perceptron (MLP) and advanced deep learning algorithms, such as LSTM and CNN, were applied to forecast neonatal mortality in Ethiopia. All neural networks were built using the Keras deep-learning framework in Python.

Data Preparation

The time-series dataset was split into 2 subsets, training and test datasets, whereby 68% of the dataset was used for training and the remaining 32% was used to test the accuracy of the models. Unlike the standard train-test split technique, which randomly splits the data, the temporality of the dataset should be maintained in time-series data partitioning. Hence, the first 21 years of data (from 1990 to 2010) were used for training, and the remaining 10 years of data (from 2011 to 2020) were used to test model performance.

As this study used univariate time-series data, the sequential nature of the data required preparation before it could be used to train a supervised learning model such as neural networks. Hence, the training data were prepared in tabular form and models were then developed. A supervised learning algorithm requires data to be provided as a collection of samples, where each sample has an input component (X) and output component (y). The number of time lags used as the independent variable (X) was determined by trying the number of time lags (n_input) and the time lag that minimized the forecasting error. The results of the hyperparameter tuning revealed that the best-performing model was obtained when the previous 4 observations (n_input=4) were used as independent variables to predict the fifth observation as the dependent variable. This means that the data from 1990 to 1993 was used as an input variable to predict the 1994 neonatal mortality rate, and the 1991 to 1994 rate was used to predict the neonatal mortality rate of 1995, and so on.

Forecasting and Model Performance Assessment

Time-series forecasting models can be evaluated on a test set using walk-forward validation, an approach in which the model makes a forecast for each observation in the test set, one at a time. After each forecast was made for a time step in the test dataset, true observations for the forecast were added to the test dataset and made available to the model. Nevertheless, the true observation for the time step could then be used as part of the input to make a prediction for the next time step. Using this forecasting approach, 10 years' worth of data (from 2011 to 2020) was predicted, the predictions were compared to the true values in the test set, and an error measure was calculated.

XSL•FO
RenderX

To evaluate the predictive performance of the forecasting method, commonly applicable measures, such as R^2, root mean squared error (RMSE), and mean absolute percentage error (MAPE), were used. Accordingly, the mode that maximizes R^2 or minimizes RMSE or MAPE is selected as the best model. Finally, the best model was used to forecast the neonatal mortality rate for the next 10 years, from 2021 to 2030, with a 95% prediction interval (PI). Figure 1 shows the overall workflow of data preparation and analysis.

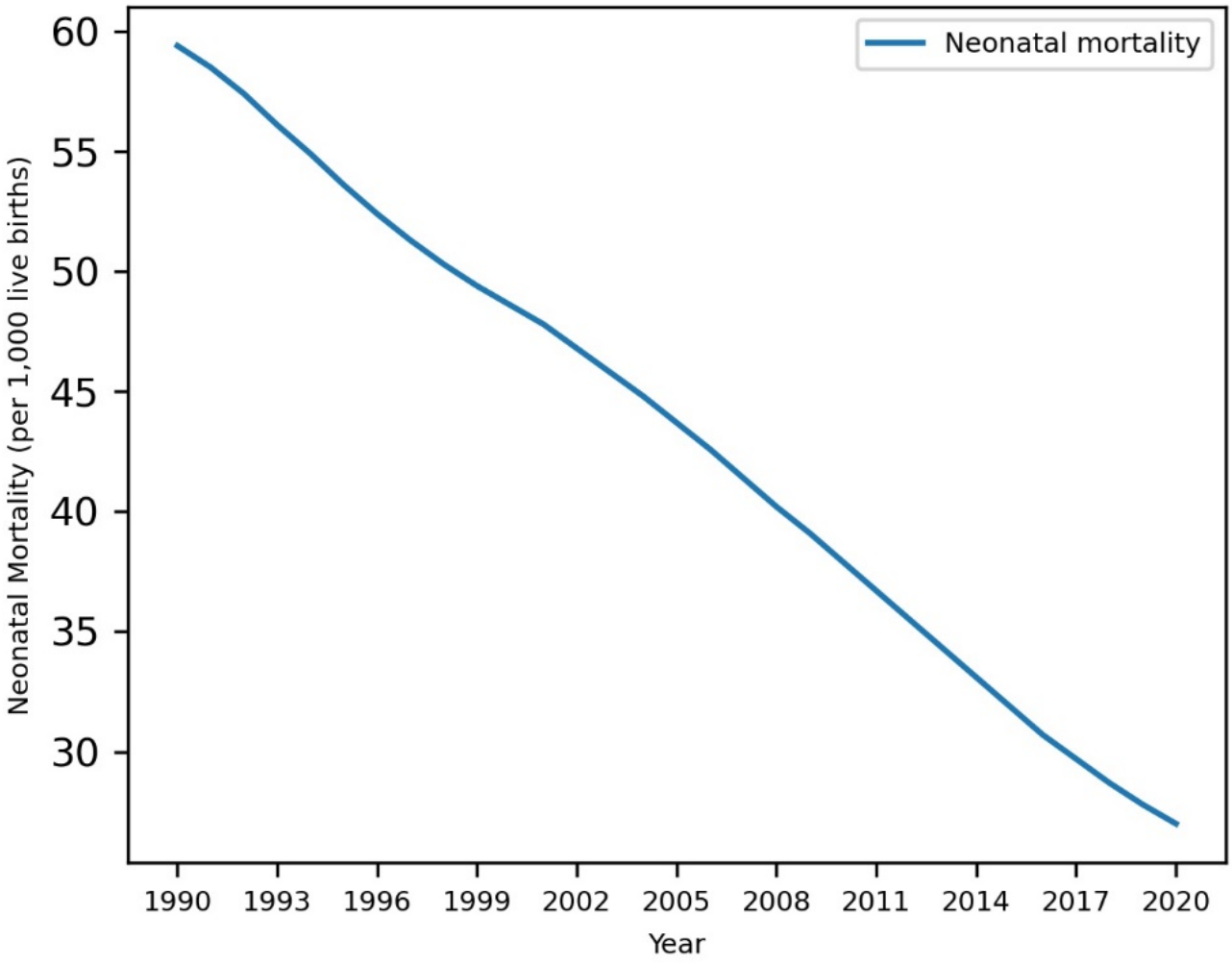
Ethical Considerations

This study used fully anonymized, aggregate-level data from the publicly accessible World Bank Health, Nutrition, and Population Statistics database. While there are no publicly available national or institutional policies in Ethiopia that explicitly address the ethics review requirements for such publicly available secondary data use, the nature of the data and the research process met the exemption criteria for ethics review under internationally recognized standards. Per the Government of Canada Panel on Research Ethics, as this work involved analysis of preexisting, publicly disseminated data, it did not involve any interaction with human participants and presented zero reidentification risk due to the aggregated nature of the data, this study did not require research ethics board review or approval per the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2 - 2022, Articles 2.2 and 2.4) [16]. Therefore, the study was conducted in compliance with recognized ethical standards for the use of public secondary data.

Results

Neonatal mortality has declined significantly in Ethiopia, from 59.4 per 1000 live births in 1990 to 27 per 1000 live births in 2020. Figure 2 shows that neonatal mortality has shown a decreasing trend over time. This trend was eliminated by first-order differencing as the best order in the hyperparameter tuning result; hence, neonatal mortality after differencing was stationary and ready for forecasting.

Figure 2. Neonatal mortality in Ethiopia (1990 - 2020).



Hyperparameter Tuning

For each model applied in the analysis, grid search hyperparameter tuning was used to develop the best model with possible optimal values. Table 1 shows the hyperparameters tried and the optimal values that minimized the forecasting errors.

XSL•FO
RenderX

Table . Results of model hyperparameter tuning.

Model	Model hyperparameters[a]	Hyperparameter space[b]	Total combination	Best combination
ARIMA[c]	• p, d, q	• q=[0,1,2,3] • d=[0,1,2,3] • p=[0,1,2,3,4]	80	• p=2 • d=1 • q=2
Double ES[d]	• trend • damped	• trend=['add', 'mul'] • damped=[True, False]	4	• trend= 'add' • damped=True
MLP[e]	• n_input • n_nodes • n_batch • n_diff	• n_input = [2-5] • n_nodes = [100, 150] • n_batch = [1, 150] • n_diff = [0,1, 2,4]	64	• n_input=4 • n_nodes=100 • n_batch=1 • n_diff=1
LSTM[f]	• n_input • n_nodes • n_batch • n_diff	• n_input=[2-5] • n_nodes=[100, 150] • n_batch=[1, 150] • n_diff=[0,1, 2,4]	64	• n_input=4 • n_nodes=100 • n_batch=1 • n_diff=1
CNN[g]	• n_input • n_diff	• n_input=[2-5] • n_diff=[0,1, 2,4]	16	• n_input=4 • n_diff=1

[a]Hyperparameters are model-specific settings tuned to optimize performance.

[b]Values in brackets represent the range or options tested using grid search.

[c]ARIMA: autoregressive integrated moving average.

[d]ES: exponential smoothing.

[e]MLP: multilayer perceptron.

[f]LSTM: long short-term memory.

[g]CNN: convolutional neural network.

The hyperparameters of all the models were tuned in a grid search manner, and the best combination of hyperparameters was used to develop the final model. Accordingly, the lag order (p), degree of differencing (d), and order of the moving average (q) are the hyperparameters of ARIMA. Similarly, the trend and damped hyperparameters of double exponential smoothing were tuned.

The number of previous inputs to be used as inputs for the model (n input), the number of nodes to be used in the hidden layer (n_nodes), and the number of samples to be included in each mini-batch (n_ batch) were the tuned hyperparameters for neural networks (ie, MLP, LSTM, and CNN). Furthermore, the difference order (n_diff) was tuned to obtain the optimal value for removing the trend from the data. The n_diff optimal value for all neural networks is 1, suggesting that the value of a previous observation should be subtracted from the current observation before fitting the model and that the predictions of the model need to be reversed to obtain the normal forecasted data. Table 1 presents the full range of hyperparameters tested for each model (the hyperparameter space), the total number of combinations evaluated, and the best-performing configuration based on model evaluation criteria (eg, lowest RMSE or MAE). The hyperparameters listed in the table correspond to parameters in the Python libraries used in this study: *statsmodels* for classical models (ARIMA and Exponential Smoothing) and *Keras* for deep learning models (MLP, LSTM, and CNN).

Model Comparison

After developing the final models based on the optimal hyperparameters of the training data, the performance of the fitted models on the test data was assessed, as presented in Table 2. The results showed that the double exponential smoothing model was the best, with a maximum R^2 of 99.94% and minimum MAPE and RMSE of 0.0015 and 0.0748, respectively. The worst-performing among the five models was the CNN, with an R^2 of 93.71% and a maximum RMSE of 0.7903.

This performance was also clearly observed graphically, with an overlap of the predicted and actual data in double exponential smoothing and a large gap in the CNN model (see Figure 3).

XSL•FO
RenderX

Table . Model performance comparison results.

Model	R^2	MAPE[a]	RMSE[b]
Double exponential smoothing	0.9994	0.0015	0.0748
ARIMA[c]	0.9992	0.0025	0.0864
MLP[d]	0.9884	0.0094	0.3396
LSTM[e]	0.9830	0.0116	0.4110
CNN[f]	0.9371	0.0237	0.7903

[a]MAPE: mean absolute percentage error.

[b]RMSE: root mean squared error.

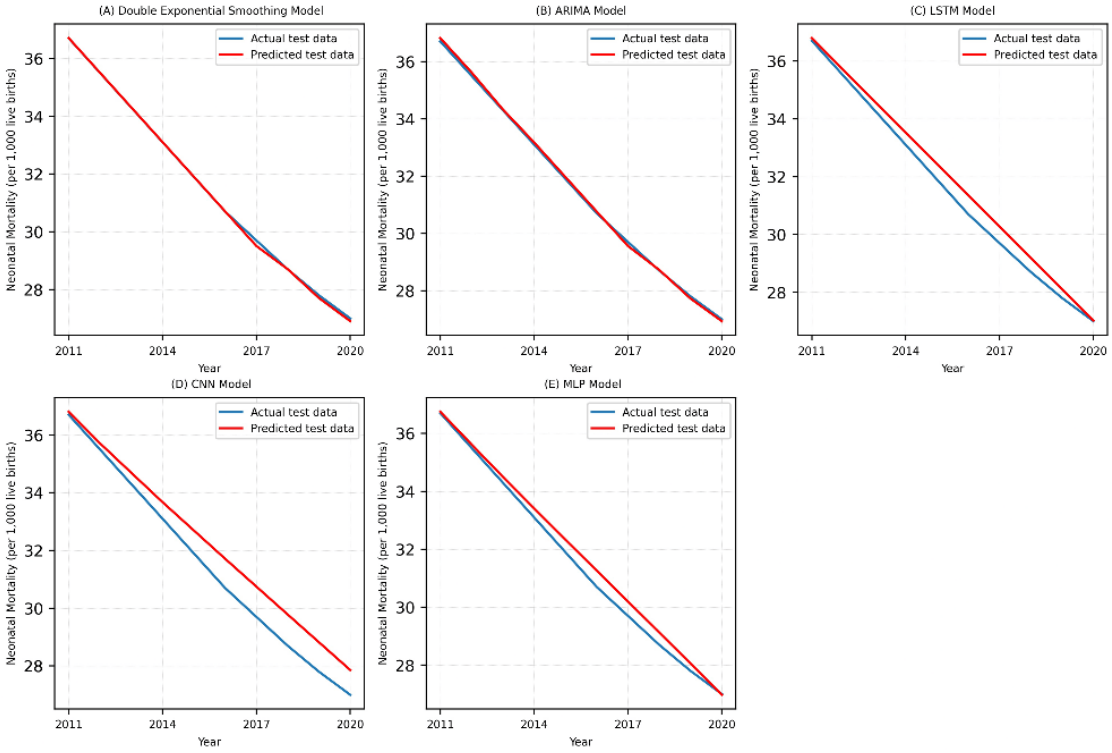
[c]ARIMA: autoregressive integrated moving average.

[d]MLP: multilayer perceptron.

[e]LSTM: long short-term memory.

[f]CNN: convolutional neural network.

Figure 3. Performance comparison of classical and deep learning models on the test data. ARIMA: autoregressive integrated moving average; LSTM: long short-term memory; CNN: convolutional neural network; MLP: multilayer perceptron.



Final Prediction

Neonatal mortality was predicted from 2021 to 2030 using a double exponential smoothing model with a 95% PI (see Table 3 and Figure 4). Accordingly, neonatal mortality in Ethiopia was forecasted to be 23.2 (PI 22.2-24.4) per 1000 live births in 2025 and 19.8 (PI 17.1-22.8) per 1000 live births in 2030. From this estimation, it is clear that Ethiopia is slightly behind what is necessary to achieve the HSTP-II neonatal mortality rate target of 21 per 1000 live births by 2025 and is also highly unlikely to meet the SDG target of 12 or fewer neonatal deaths per 1000 live births by 2030.

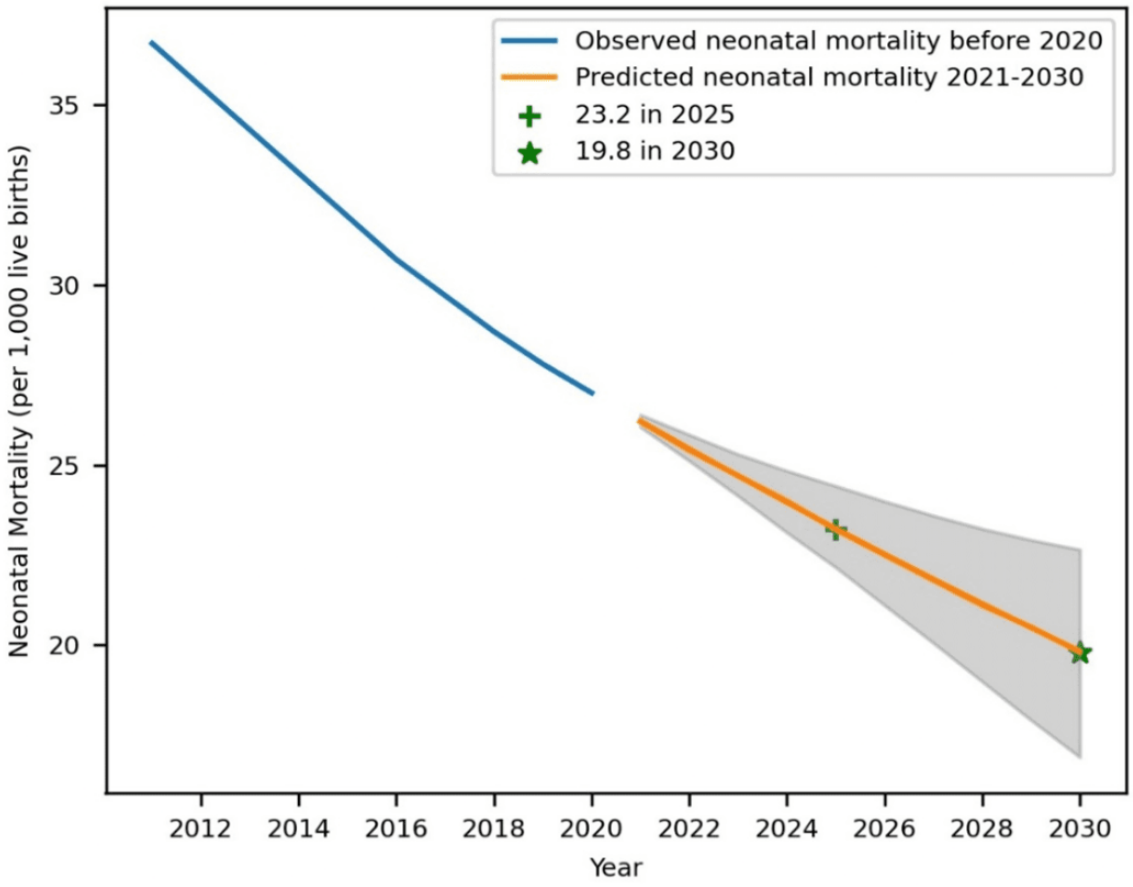
XSL•FO
RenderX

Table . Neonatal mortality forecasting in Ethiopia (2021 to 2030).

Year	Predicted neonatal mortality, deaths per 1000 live births (PI[a])
2021	26.2 (26.1-26.4)
2022	25.4 (25.1-25.8)
2023	24.7 (24.2-25.2)
2024	24.0 (23.2-24.8)
2025	23.2 (22.2-24.4)
2026	22.5 (21.2-23.9)
2027	21.8 (20.2-23.6)
2028	21.1 (19.1-23.4)
2029	20.5 (18.1-23.1)
2030	19.8 (17.1-22.8)

[a]PI: prediction interval.

Figure 4. Prediction of neonatal mortality in Ethiopia from 2021 to 2030.



Discussion

Principal Findings

In this study, the performance of deep-learning and classical time-series models in predicting neonatal mortality rates in Ethiopia was compared. The methods included ARIMA, exponential smoothing, MLP, CNN, and LSTM. The comparison was based on evaluation metrics such as R^2, MAPE, and RMSE. Based on these results, exponential smoothing performed best, with the lowest MAPE and RMSE values of 0.0015 and 0.0748, respectively. ARIMA was also the best-performing model, followed by exponential smoothing, for predicting neonatal mortality. However, the CNN performed the worst among the 5 models. This result is compatible with previous studies that reported that classical methods are more accurate than neural network models [17-19]. This result is also in line with a study that reported that LSTMs are easily outperformed by simpler methods and may not be suitable for simple autoregressive-type univariate time-series forecasting tasks [20]. In a study aimed

XSL•FO
RenderX

at clearly demonstrating the capability of a suite of different machine-learning methods compared to classical time-series forecasting methods evaluated across a collection of 1045 monthly time series, simple classical methods such as exponential smoothing and ARIMA models outperformed the machine-learning methods, including 2 modern neural network algorithms, such as recurrent neural networks and LSTM [21]. This may be owing to the appropriateness of classical methods in the case of univariate small datasets, where deep-learning algorithms are not yet at their best [22].

Another aim of this study was to forecast neonatal mortality to verify whether Ethiopia has achieved national and international targets. The results revealed that neonatal mortality in Ethiopia will be 23.2 (PI 22.2-24.4) per 1000 live births in 2025 and 19.8 (PI 17.1-22.8) per 1000 live births in 2030. These results reveal Ethiopia's inability to achieve the SDG of reducing neonatal mortality, possibly because the COVID-19 pandemic and the armed conflict that occurred in 2019 jeopardized the health care system and made it difficult for women to receive prenatal and postnatal care due to lockdowns and restrictions, which in turn created vulnerable populations, exposing expectant mothers and newborns to danger. Furthermore, ongoing conflict in some areas of Ethiopia means that some health care facilities are nonoperational or services are interrupted, which consequently makes it difficult to achieve the SDG. The risk of neonatal infections and malnutrition may further increase in conflict zones due to inadequate sanitization, food insecurity, and access to clean water. The ability to mitigate these issues and create stable environments, which is needed to make significant advancements in lowering neonatal mortality in Ethiopia, depends on the peaceful resolution of conflicts and humanitarian aid.

Limitations

One of the primary limitations of this study is the relatively small size of the dataset, which includes annual neonatal mortality rates from 1990 to 2020. While this limited data span presents challenges, such as an increased risk of model overfitting and higher prediction uncertainty, we have taken extensive measures to address these issues. Specifically, we have applied advanced optimization techniques, including rigorous regularization, hyperparameter tuning, and cross-validation, to ensure that our models are robust and effectively fit the data. These efforts help to mitigate the impact of the small dataset on the accuracy and reliability of our forecasts for the period from 2021 to 2030. Although the dataset size constrains the breadth of variability and external factors captured, our careful optimization and validation processes enhance the credibility of the results. Future research could build on this foundation by incorporating larger and more diverse datasets, which would further strengthen the robustness of predictive models and provide even more accurate forecasts.

Conclusion

This study found that classical models outperformed advanced neural network models in forecasting neonatal mortality in Ethiopia. This implies the need to use classical time-series methods in time-series forecasting rather than relying on advanced deep-learning models. This study also revealed that the national (HSTP-II) and international (SDG) targets for neonatal mortality cannot be realized if the current trend continues. This highlights that urgent interventions are necessary to strengthen the health care system and accelerate the decline in the neonatal mortality rate. To achieve these targets, collaborative efforts with concerned stakeholders are also needed to improve the health care system and provide responsive neonatal and child health services.

Acknowledgments

The author is grateful to the World Bank for making the data publicly available. The author declared that he had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided APF support for the publication of this paper.

Data Availability

The datasets generated and analyzed during this study are publicly available at the World Bank data repository and can be accessed at [23].

Authors' Contributions

SDK contributed to conceptualization, methodology, data curation, software, formal analysis, writing the original draft, validation, and writing – review and editing.

Conflicts of Interest

None declared.

References

1. Neonatal mortality rate (per 1000 live births). World Health Organization. 2021. URL: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/67 [accessed 2025-07-24]

XSL•FO
RenderX

2. Rosa-Mangeret F, Benski AC, Golaz A, et al. 2.5 million annual deaths-are neonates in low- and middle-income countries too small to be seen? A bottom-up overview on neonatal morbi-mortality. Trop Med Infect Dis 2022 Apr 21;7(5):64. [doi: 10.3390/tropicalmed7050064] [Medline: 35622691]
3. Levels & trends in child mortality: report 2020, estimates developed by the united nations inter-agency group for child mortality estimation. : United Nations Inter-agency Group for Child Mortality Estimation (UN IGME); 2020 URL: https://www.unicef.org/media/79371/file/UN-IGME-child-mortality-report-2020.pdf [accessed 2025-07-24]
4. Doku DT. Survival analysis of neonatal mortality in Ghana using three population-based surveys. Arch Public Health 2022 Jan 7;80(1):21. [doi: 10.1186/s13690-021-00773-3] [Medline: 34996517]
5. United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). Trends in child mortality: report 2018. 2018 URL: https://data.unicef.org/wp-content/uploads/2018/10/Child-Mortality-Report-2018.pdf [accessed 2025-07-29]
6. Ethiopia demographic and health survey 2016: key indicators report. : ICF; 2016 URL: https://www.cmpethiopia.org/media/ethiopia_demographic_health_survey_2016_key_indicators [accessed 2025-07-29]
7. Sachs JD, et al. Sustainable Development Report 2022: Cambridge University Press; 2022. [doi: 10.1017/9781009210058]
8. Health sector transformation plan II (HSTP II) 2020/21–2024/25. : Ministry of Health Ethiopia; 2021 URL: http://repository.iphce.org/bitstream/handle/123456789/3742/HSTP-II-MTR-final-report-2023.pdf [accessed 2025-07-29]
9. Zeroual A, Harrou F, Dairi A, Sun Y. Deep learning methods for forecasting COVID-19 time-series data: a comparative study. Chaos, Solitons & Fractals 2020 Nov;140:110121. [doi: 10.1016/j.chaos.2020.110121]
10. Health nutrition and population statistics. World Bank Group. URL: https://databank.worldbank.org/source/health-nutrition-and-population-statistics [accessed 2025-07-24]
11. Child mortality estimates. UN Inter-agency Group for Child Mortality Estimation. URL: https://childmortality.org/ [accessed 2025-07-24]
12. Kotu V, Deshpande B. Time series forecasting. In: Kotu V, Deshpande B, editors. Data Science, 2nd edition: Morgan Kaufmann; 2019:395-445. [doi: 10.1016/B978-0-12-814761-0.00012-5]
13. Siami-Namini S, Tavakoli N, Siami Namin A. A comparison of ARIMA and LSTM in forecasting time series. Presented at: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA); Dec 17-20, 2018; Orlando, FL, USA. [doi: 10.1109/ICMLA.2018.00227]
14. Brownlee J. Deep Learning for Time Series Forecasting: Predict the Future With MLPs, CNNs and LSTMs in Python: Machine Learning Mastery; 2018.
15. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice: OTexts; 2018.
16. Panel on research ethics: TCPS 2 (2022) – Chapter 2: scope and approach. Government of Canada. 2022. URL: https://ethics.gc.ca/eng/tcps2-eptc2_2022_chapter2-chapitre2.html [accessed 2025-07-09]
17. Choudhary M, Jain S, Arya G. Classical Models vs Deep Leaning: Time Series Analysis: Springer Nature Switzerland; 2022. [doi: 10.1007/978-3-031-23724-9_26]
18. Jouilil Y, Iaousse M. Comparing the accuracy of classical and machine learning methods in time series forecasting: a case study of USA inflation. Stat, optim inf comput 2023;11(4):1041-1050 [FREE Full text] [doi: 10.19139/soic-2310-5070-1767]
19. Uzzal MH, Ślepaczuk R. The performance of time series forecasting based on classical and machine learning methods for S&P 500 index. University of Warsaw. 2023. URL: https://www.wne.uw.edu.pl/application/files/4016/7775/9406/WNE_WP412.pdf [accessed 2025-07-29]
20. Gers FA, Eck D, Schmidhuber J. Applying LSTM to time series predictable through time-window approaches. In: Artificial Neural Networks - ICANN 2001: Springer; 2001. [doi: 10.1007/3-540-44668-0_93]
21. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: concerns and ways forward. PLoS ONE 2018;13(3):e0194889. [doi: 10.1371/journal.pone.0194889] [Medline: 29584784]
22. Iaousse M, Jouilil Y, Bouincha M, Mentagui D. Comparative simulation study of classical and machine learning techniques for forecasting time series data. Int J Onl Eng 2023;19(8):56-65. [doi: 10.3991/ijoe.v19i08.39853]
23. World development indicators. World Bank Group. URL: https://databank.worldbank.org/reports.aspx?source=2&series=SH.TBS.INCD&country=ETH&_gl=1*bw5pab*_gcl_au*MTI2Mjk3MzY2MC4xNzI0MjcwMDc3 [accessed 2025-07-24]

Abbreviations

ACF: autocorrelation function
ARIMA: autoregressive integrated moving average
CNN: convolutional neural network
HSTP-II: second Ethiopian Health Sector Transformation Plan
LSTM: long short-term memory
MA: moving average
MAPE: mean absolute percentage error
MDG: Millennium Development Goals
MLP: multilayer perceptron
PACF: partial autocorrelation function

XSL•FO
RenderX

PI: prediction interval
RMSE: root mean squared error
SDG: Sustainable Development Goal
UN IGME: United Nations Inter-agency Group for Child Mortality Estimation
UNICEF: United Nations International Children's Emergency Fund

Edited by E Mensah; submitted 23.09.24; peer-reviewed by T Okui; revised version received 11.05.25; accepted 20.06.25; published 25.08.25.

Please cite as:
Kebede SD
Forecasting Neonatal Mortality in Ethiopia to Assess Progress Toward National and International Reduction Targets Using Classical Techniques and Deep Learning: Time-Series Forecasting Study
Online J Public Health Inform 2025;17:e66798
URL: https://ojphi.jmir.org/2025/1/e66798
doi:10.2196/66798

© Shimels Derso Kebede. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 25.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Patient Perceptions of Bacterial Vaginosis Treatments: Mixed Methods Sentiment Analysis Study of Online Drug Review Forums

Eren Watkins[1], PhD, MPH; Andrea N Cimino[2], PhD, MSW; Curtis Culbertson[2], MS; Juliana Raymaker[1], MD; Jennifer R Amico[3], MD, MPH

[1]Medical Affairs & Outcomes Research, Organon & Co, Jersey City, NJ, United States

[2]Rogue Scholar Consulting, Baltimore, MD, United States

[3]Family Medicine and Community Health, Robert Wood Johnson Medical School, Rutgers Health, New Brunswick, NJ, United States

Corresponding Author:
Eren Watkins, PhD, MPH
Medical Affairs & Outcomes Research
Organon & Co
30 Hudson St
Jersey City, NJ, 07302
United States
Phone: 1 404 670 4920
Email: eren.watkins@organon.com

Abstract

Background: Bacterial vaginosis (BV) is the most common cause of vaginal discharge in people of childbearing age in the United States. More information about what patients do and do not like about the most common BV products and the extent to which they reduce BV symptoms is important for understanding patients' health and the current treatment landscape for BV.

Objective: Using data from online drug review forums, this study's objectives were to (1) quantitatively characterize the patient voice via sentiments (positive to negative) and emotions about the three most common Food and Drug Administration (FDA)–approved treatments for BV—oral metronidazole (OM), vaginal metronidazole (VM), vaginal clindamycin (VC)—and (2) qualitatively summarize themes characterizing the patient-perceived impact of BV and BV products.

Methods: Data for this mixed methods descriptive study came from 1645 users' reviews of BV products posted on WebMD.com and Drugs.com. Reviewer attributes, reviewer-submitted star ratings, and sentiment analysis (SA) using word-emotion association were analyzed with descriptive statistics and bivariate associations. A traditional qualitative analysis using qualitative description was also performed.

Results: Most reviewers were female (n=629, 99.4%), between the ages of 18 and 44 years, and reported using BV products for less than 1 month, though qualitative results suggested most reported recurrent BV infections. Quantitative results revealed reviewers' preference for vaginal products. The mean star ratings for VC were significantly higher when compared to OM and VM. VC reviews had the highest proportion of positive emotion words compared to OM and VM. Qualitative results for VC supported the quantitative findings: favorable themes related to perceptions of value, effectiveness in alleviating symptoms, and minimal side effects. Additionally, despite some concerns related to the cost of VC, reviewers said they would use the medication again. Other qualitative findings supported BV medical education campaigns for patients and providers on BV treatment.

Conclusions: Overall, people want a BV treatment that is easy to use, quickly alleviates symptoms, and has minimal side effects. Patients use product reviews to inform their decision-making about BV treatment, ask and seek answers to health-related questions, and share their experiences, presenting a unique opportunity for comprehensive patient education through clinical encounters or public health outreach efforts.

(Online J Public Health Inform 2025;17:e71720) doi:10.2196/71720

KEYWORDS

bacterial vaginosis; bacterial vaginosis symptoms; patient experiences; bacterial vaginosis treatment; sentiment analysis; natural language processing; qualitative

XSL•FO
RenderX

Introduction

Background

Bacterial vaginosis (BV) is the most common cause of vaginal discharge in people of childbearing age in the United States [1], with approximately one-third (29%) of those aged 14-49 years having experienced either symptomatic or asymptomatic BV [2]. BV is caused by an overgrowth of anaerobic bacteria naturally found in the vagina, causing symptoms such as itching, a "fishy" odor, burning while urinating, and changes in the color and consistency of vaginal discharge [2]. BV acquisition is associated with myriad factors, including sexual health and hygiene practices (eg, vaginal douching, having multiple male or female sexual partners) [3,4], stress exposure, smoking, and being African American [4-6]. BV is associated with increased susceptibility to sexually transmitted infections (STIs), including HIV [7]; gynecological conditions, such as pelvic inflammatory disease (PID) and endometritis [3]; and adverse pregnancy outcomes, such as pregnancy loss and miscarriage, preterm labor, and low birth weight [3,8,9].

Several prescription medications (eg, nitroimidazole, clindamycin) and over-the-counter (OTC) products, such as boric acid, are commonly used to treat BV [10]. Treatment selection is often made based on availability, side effects, and cost [11]. Many health care providers who are satisfied with the status quo have developed the practice of routinely choosing from among a few "tried and true" (eg, oral metronidazole [OM]) [12] treatments that are easily available, inexpensive, and familiar. For example, in a recently published study to estimate the frequency of BV recurrence and treatment patterns among 140,826 commercially insured patients, the most commonly prescribed BV medication was OM (73.6%) [13]. Although BV treatments result in 80% and 90% cure rates [14], more than 50% of women experience BV reinfection within months of initial treatment [15,16].

The internet is a key information source for patients with BV infection [17], particularly given the stigma and negative impact on the quality of life associated with BV and its symptoms [18,19]. Online reviews of BV products can be a powerful tool patients use to evaluate perceived effectiveness, side effects, and other information that may ultimately influence their decision to use a product. These unsolicited, unfiltered product reviews are a novel data source to support real-world evidence generation of the patient voice. Even Food and Drug Administration (FDA) [20] guidance recognizes the potential of online tools as useful data sources "to complement literature review findings, inform the development of research tools (e.g., qualitative study discussion guides), or as a supplement to traditional research approaches (e.g., literature, one-on-one interviews, focus groups, or expert opinion)." Understanding what people do and do not like about these products and the extent to which a product reduces their BV symptoms is an important step to understanding patients' health and the current treatment landscape for BV.

Sentiment Analysis to Uncover Opinions

Unsolicited, unfiltered product reviews are a novel data source to support real-world evidence generation of the patient voice. However, mining online product reviews yields enormous amounts and types (structured vs unstructured) of passively collected data that are challenging to analyze. Sentiment analysis (SA), otherwise known as opinion mining, is a type of natural language processing that calculates the degree of positive, negative, or neutral sentiments based on words or phrases in a text [21]. SA uses various algorithms to assign a numerical score to precompiled lexicons or dictionaries of words or phrases previously coded in linguistic research. For example, the *Word-Emotion Association Lexicon* created by Mohammad and Turney [22] at the National Research Council Canada (NRC) contains 13,889 words and has been used to analyze sentiment and emotions in words, sentences, tweets, and entire documents from a variety of fields, including psychology, behavioral and social sciences, and psycholinguistics. SA algorithms can be complex, leveraging sentence structure or clauses with contrasting sentiments to assign weights, or they can be straightforward counts of positive and negative words referenced in a lexicon. The benefit of using SA is that it can quickly categorize reviewer sentiment about a product and provides a common metric to compare sentiments across products.

The purpose of this study was to gain a better understanding of patient-reported sentiment toward BV medications from user reviews of BV products identified via online (public) drug review forums. The study focused on three FDA guideline–recommended BV treatments: OM, vaginal metronidazole (VM), and vaginal clindamycin (VC). The objectives of the study were to (1) quantitatively characterize patient-reported star ratings, sentiments (polarity of positive to negative), and emotions about BV and BV products from patients who had submitted product reviews and (2) qualitatively summarize themes characterizing the patient-reported impact of each BV product on the quality of life and health care provider interactions.

Methods

Study Design

This mixed methods, descriptive study examined BV product reviews from Drugs.com and WebMD.com for OM, VM, and VC. We selected these websites because of their ubiquitous popularity. WebMD.com is "the best known and most popular consumer-oriented health site" [23], whereas Drugs.com "has more than 50 million users each month, most of whom are in the US" and provides its users with "trusted, accurate, and ethical information about thousands of medications" [24].

Ethical Considerations

The study protocol was approved by Solutions IRB (protocol #2022/12/19), an independent institutional review board service, as nonhuman subject research because the data were publicly available for secondary analysis, included no identifiable information, and involved no interaction between investigators and participants. An additional review from the study's sponsor included timely reporting of pharmacovigilance and resolution of ethical concerns.

XSL•FO
RenderX

Sample of Extracted Reviews

A custom web scraper was developed to automate downloading and organizing of the data into a relational Structured Query Language (SQL) database for analysis in R (R Foundation for Statistical Computing). Data included reviews for OM, VM, and VC used for "bacterial vaginosis" or "vaginosis caused by bacteria," as selected by the reviewer upon posting a review. Eligible reviews were (1) in English, (2) from the United States, and (3) beginning with the earliest review available from September 25, 2007, through September 28, 2023. Reviews were ineligible if the reviewer used the product for a non-BV condition (ie, yeast infection, urinary tract infection, *Gardnerella* infection, trichomoniasis). A SQL query aided in identifying these ineligible reviews, which were manually verified before excluding. Of 1733 reviews that were scraped, 88 (5.1%) were ineligible, leaving a total sample of 1645 (94.9%) reviews.

Procedure

Data Cleaning

Extracted reviews were first examined to determine whether they contained reportable adverse events (AEs). No reportable AEs were detected. Next, we examined the authenticity of the reviews to ensure a bot was not repeatedly posting the same review. Thus, we deployed a similarity analysis to look for reviews with >5 words that matched another extracted review by 90% or more. One pair of reviews for OM showed the same reviewer posted a mostly similar review to both WebMD.com and Drugs.com; therefore, these reviews were deemed real and were not excluded from the extracted sample. Next, we followed industry-standard steps [22] to prepare the data for SA, including (1) data cleaning (ie, converting text to lowercase, removing/replacing special characters, translating emoticons into their Unicode names/descriptions, expanding contractions, and correcting typos), (2) word-level tokenization (ie, splitting up into machine-readable units), and (3) word-level standardization ("lemmatization") to reduce a word into its root or base form. Original and transformed reviews were retained in the SQL database.

SA was conducted using the NRC's *Word-Emotion Association Lexicon* [22]. In addition to positive and negative sentiments, this first-of-its-kind lexicon further categorizes words to one or more of eight universal emotions based on Plutchik's Wheel of Emotions [25,26]: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. The wheel offers a structured and nuanced way to visualize emotions and their variation in intensity (eg, low anger is annoyance, high anger is rage) and in combinations (eg, joy + trust = love). The NRC uses a count of positive words and negative words, which can be summed to derive a single score or presented as a proportion by dividing the sum by the total number of words in a corpus, such as all reviews for a product. Words can be associated with multiple emotions. For instance, the NRC lexicon classifies the word "smell" as negative and links it to both anger and disgust emotions.

Narrative Topic Classification

Individual reviews typically contained multiple topics—for instance, how BV affected the patient's health and relationships, side effects of products tried in the past, and the value of the reviewed product—which introduced random variability, or noise, in the SA scores. For instance, a reviewer who gave a 5-star review may paradoxically have a negative sentiment score if they discuss a life-long struggle with BV and a poor interaction with a health care provider. Thus, it was necessary to determine the variety of topics discussed by the reviewer. These classifications generally followed nine topic areas described in Table 1.

XSL•FO
RenderX

Table 1. Topic classifications and definitions.

Code	Definition
Value	Positive, negative ("did not work") and neutral valuation ("no side effects") statements by reviewers about the product's efficacy, price, or worth and whether they would use the product again
Side effects	Any intended or unintended negative effects of the reviewed product
Adherence	Whether reviewers stopped taking the product and why
Product characteristics	Features of the product, such as attributes, taste, consistency, method of delivery, and packaging
Recommended use	Statements about how to use the product and minimize side effects or tips to more effectively manage BV[a] symptoms
Cross product	Statements discussing the effectiveness or experience using another FDA[b] or OTC[c] product to treat BV
Condition	Statements about the disease state (symptoms, triggers, recurrence, impact on life or relationships, etc)
Doctor interaction	Statements describing an interaction or conversation with a doctor, pharmacist, or other health care professional
Miscellaneous	All other statements, such as treatment plan/dosage, reviewer questions, words of encouragement, myths about BV, ways to improve the product, and other topics

[a]BV: bacterial vaginosis.

[b]FDA: Food and Drug Administration.

[c]OTC: over the counter.

The classification of topics was derived inductively through open coding of a sample of reviews by one author (AC), who then developed a codebook for deductive application to the remaining reviews [27]. Three authors (AC, EW, and SW) divided the reviews for individual coding. A second author verified all classifications to ensure consistency in coding across the dataset. A review could contain multiple classifications; however, each narrative segment (ie, a string of words) was coded with a single classification.

The topic classification informed both quantitative and qualitative analyses. Classifications related to the reviewed product's value, side effects, and adherence challenges were most relevant to the quantitative analysis of star ratings and SA scores, while those reflecting patient struggles with BV, poor doctor interactions, or dissatisfaction with prior treatments contributed primarily to the qualitative analysis.

Key Variables

Product Information

The product name, formulary, and route of administration (oral vs vaginal) were extracted from both websites (all categorical).

Reviewer Attributes

All reviewer attributes were categorical. WebMD.com collected the most reviewer information, including patient gender (male, female, nonbinary), age range (from 0-2 years to 75 years and over), and time on medication (from <1 month to 10 years or more). Gender was missing for 13.6% of reviews, the age range was missing for 3.4% of reviews, and time on medication was missing for 7.4% of reviews on WebMD.com. Drugs.com only collected information on time on medication (from <1 month to 10 years or more), and 45.7% of reviews were missing these data.

Star ratings were a primary dependent variable (continuous), with higher values indicating a more positive rating. The same product was reviewed on both Drugs.com and WebMD.com. Drugs.com asked a single question: "On a scale from 1 to 10, select how effective you found this medication, taking into consideration the benefits, side effects, and ease of use, from 1 (not effective) to 10 (most effective)." WebMD.com asked reviewers to rate the product on three separate attributes—overall satisfaction, effectiveness, and ease of use—on a scale ranging from 1 (very poor) to 5 (excellent). First, we standardized the scale by averaging the three items (min=0.5, max=5). Next, we examined whether there were statistically significant differences for the same product by website using independent *t* tests, and none were found. Because the scores did not differ by website, we averaged the star ratings for each product. Star rating scores were created for the entire review ("overall") and by classification.

NRC emotion proportions were used to characterize all words for an entire product associated with eight emotions. Emotion proportions were used to compare emotions across products because a net count would bias reviews that contained more words. The following formula was used:



Here,  is the sum of the occurrences of words associated with one of the eight emotions (where n is the total number of emotion-associated words), and  is the sum of all words in

XSL•FO
RenderX

the product (where m is the total number of words in the product text).

Classification Proportions

We quantified the proportion of reviews that mentioned each of the nine classifications as they represent the variety of topics mentioned for a given product. The following formula was used:



Statistical Analysis

Quantitative data were summarized with descriptive statistics to characterize the sample of reviews including product information, reviewer attributes, and star ratings. Next, *t* tests (two-tailed, α=.05) were used to compare star ratings by reviewer attribute. For between-product comparisons, ANOVAs with Tukey correction (two-tailed, α=.05) were used with the star rating as the dependent variable. Visualizations were a key aspect of the quantitative data analysis and enabled comparisons by product. Emotion word proportions were depicted with bar charts. Classification proportions were depicted with clustered bar charts to compare the contributions of each classification category.

Qualitative Analytic Approach

In addition to the classification coding described before, a second qualitative analysis was performed using the 40 longest reviews for each product as these were the richest cases. Many qualitative analyses require 20-30 cases to support in-depth, iterative analysis [27,28]. A sample size of 40 was selected to achieve *information power* (rather than information redundancy or saturation), a concept that suggests that the number of participants needed increases when the sample provides less information, is less relevant to the study aims, offers less depth of cases, or requires more complex analysis [29,30]. We quantified the sample's information power. Analysis showed that the selected cases contained multiple emotions. They had substantial overlap with reviews containing the highest-rated emotion scores; on average, the longest 40 reviews overlapped with 70% of the top emotion scores for VC (n=94), 56% for VM (n=195), and 32% for OM (n=1332). Lastly, the average word count in the 40 longest reviews was 139.6 (VC), 179.2 (VM), and 282.5 (OM) words.

The chosen analytic approach was qualitative description [31] because it offers a "rich, straight description of an experience or an event" [32] without the imposition of a theoretical or interpretative lens and aligned with the study aim to describe the reviewers' experience in their own language. Open and inductive coding methods were used, focused on describing patterns and themes rather than interpretive results. Only one author (AC) performed the descriptive qualitative analysis; however, themes aligned with the double-coded and verified classifications, increasing trustworthiness of the findings.

Results

Sample Description Reviewer Attributes

Of the 1645 reviews that met the eligibility criteria, 82% (n=1349) were for OM, followed by 12.3% (n=202) for VM and 5.7% (n=94) for VC (Table 2).

Nearly two-thirds of the total sample (n=993, 60.4%) was from Drugs.com. Metronidazole comprised the largest formulary (n=1551, 94.3%: oral, n=1349, 87%; vaginal, n=202, 13%). Only 18% (n=296) of reviews were for vaginal products. A majority of Drugs.com reviewers were on the medication for less than 1 month (n=499, 94.2%). Reviewer attributes from WebMD.com showed the sample was primarily female (n=629, 99.4%), between the ages of 18 and 44 years (n=574, 81%), and took the medication for less than 1 month (n=560, 92.6%). The 4 (0.6%) nonfemale cases included 1 (25%) marked as "nonbinary" and 1 (25%) "male," both of whom referenced having a vagina in their review; the other 2 (50%) cases did not contain enough review information to determine the sex assigned at birth.

Table 2. Number and breakdown of eligible reviews by product.

Product	Star rating + review text (n=1536), n (%)	Star rating + no review text (n=80), n (%)	No star rating + review text (n=100), n (%)	Total (N=1645), n (%)
VC[a]	88 (5.7)	5 (6.3)	1 (1.0)	94 (5.7)
VM[b]	189 (12.3)	8 (10.0)	5 (5.0)	202 (12.3)
OM[c]	1259 (82.0)	16 (20.0)	74 (74.0)	1349 (82.0)

[a]VC: vaginal clindamycin.

[b]VM: vaginal metronidazole.

[c]OM: oral metronidazole.

Disease State

Results describing the disease state (qualitative n=120) from the 40 longest VC, OM, and VM reviews showed that most reviewers reported suffering from recurrent BV (VM=20, OM=11, VC=17), ranging from a few months to 14 years. A portion of reviewers reported having asymptomatic BV that was diagnosed at a routine well-woman exam. More reviewers who self-reported this was their first time with BV left reviews for VM (n=10), followed by OM (n=8) and VC (n=1). The BV symptoms reported were a "fishy" smell, itching, and irritation. BV triggers included menstruation, unprotected vaginal or oral sex, ejaculation, gynecological conditions (eg, childbirth/pregnancy), intrauterine device (IUD) insertion, hysterectomy, and, to a lesser extent, douching and superabsorbent tampons.

XSL•FO
RenderX

Star Ratings

There were no statistically significant differences in star ratings by time on medication (Table 3), the only demographic characteristic collected by WedMD.com and Drugs.com. Products administered vaginally (mean 3.5, SD 1.48) had statistically significant higher star ratings compared to oral products (mean 3.31, SD 1.44, $t_{423.92}$=–2.34, 95% CI –0.41 to –0.04, P<0.05). VC had the highest star rating (mean 3.85, SD 1.32), followed by VM (mean 3.39, SD 1.52) and OM (mean 3.31, SD 1.44). This difference in star ratings between the products was statistically significant ($F_{2,1562}$=5.98, P<.01), with VC being 0.54 points higher than OM (95% CI –0.96 to –0.10, P<.01) and 0.46 points higher than VM (95% CI –0.88 to –0.03, P<.05). VM and OM were not significantly different. Finally, OM had the lowest average star rating among reviewers who made statements about side effects (mean 2.79, SD 1.46), followed by VC (mean 3.02, SD 1.53) and VM (mean 3.07, SD 1.5).

Table 3. Product star ratings for the review overall and by classification.

Classification[a]	VC[b] (n=94)		OM[c] (n=1349)		VM[c] (n=202)	
	Star rating, mean (SD)	Classifications, n	Star rating, mean (SD)	Classifications, n	Star rating, mean (SD)	Classifications, n
Overall	3.85 (1.32)	94	3.31 (1.44)	1349	3.39 (1.52)	202
Value	4.08 (1.25)	177	3.45 (1.48)	2593	3.62 (1.05)	330
Side effects	3.02 (1.53)	73	2.79 (1.46)	2142	3.07 (1.5)	262
Adherence	0.5 (0.71)	2	1.94 (1.31)	282	1.87 (1.21)	21
Product characteristics	3.98 (0.63)	12	3.48 (1.27)	525	3.63 (1.47)	31
Recommended use	3.68 (1.21)	38	3.83 (1.19)	1233	3.73 (1.58)	79
Cross product	3.67 (1.43)	49	3.03 (1.54)	247	3.17 (1.62)	86
Condition	3.75 (1.2)	73	3.39 (1.44)	1191	3.39 (1.56)	305
Doctor interaction	3.21 (1.41)	17	2.63 (1.62)	153	2.96 (1.89)	46
Miscellaneous	3.59 (1.4)	76	3.47 (1.45)	1170	3.43 (1.54)	180

[a]The number of classifications may exceed the total number of overall reviews, as a single review could contain multiple narrative segments (ie, a string of words) assigned to a classification.

[b]VC: vaginal clindamycin.

[c]OM: oral metronidazole.

[d]VM: vaginal metronidazole.

NRC Emotion Word Proportions

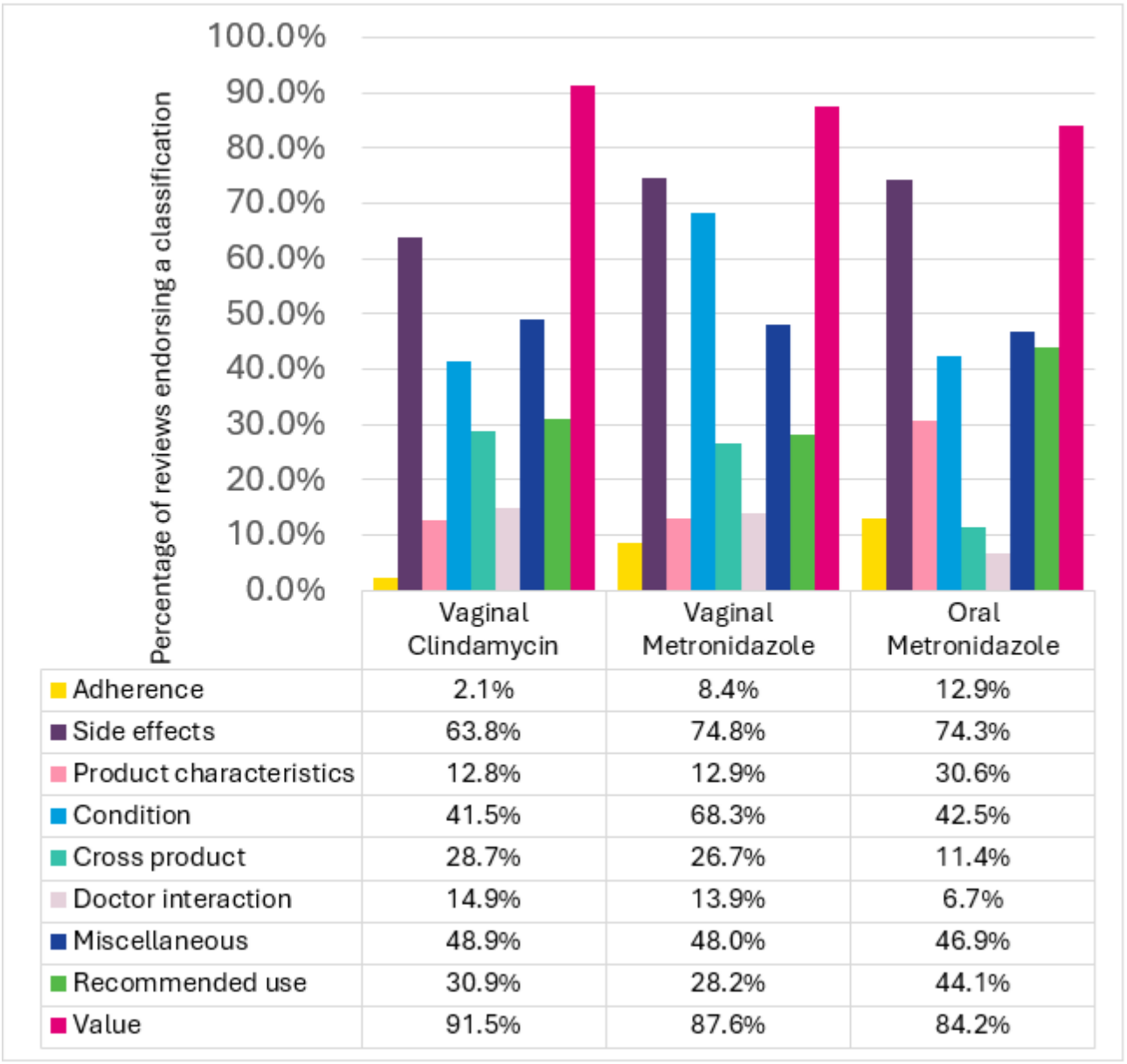
Comparing the proportion of positive and negative emotion words across products (see Multimedia Appendices 1 and 2, respectively) revealed that VC reviews had the highest proportion of three of four positive emotion words, except surprise, which was about even for VC and OM. VM and OM reviews had similar proportions of positive emotions, except for trust and surprise, which was higher for OM. In addition, OM reviews had the highest proportion of all four negative emotion words compared to VC and VM. Negative emotion word proportions for VC and VM were about the same, except for fear, which was higher for VC, and disgust, which was higher for VM.

Classification Proportions

Product classifications—themes describing the variety of topics within a single review—were depicted as a proportion of reviews for a given product (Figure 1). Value statements were the highest among VC reviews (n=86, 91.5%) compared to reviews for VM (n=177, 87.6%) and OM (n=1136, 84.2%). Difficulties in adherence (yellow) were the highest in reviews for OM (n=174, 12.9%) and the lowest in reviews for VC (n=2, 2.1%). Complaints of side effects (purple) were the lowest in reviews for VC (n=60, 63.8%) compared to reviews for OM (n=1002, 74.3%) and VM (n=151, 74.8%).

XSL•FO
RenderX

Figure 1. Classification proportions.



	Vaginal Clindamycin	Vaginal Metronidazole	Oral Metronidazole
Adherence	2.1%	8.4%	12.9%
Side effects	63.8%	74.8%	74.3%
Product characteristics	12.8%	12.9%	30.6%
Condition	41.5%	68.3%	42.5%
Cross product	28.7%	26.7%	11.4%
Doctor interaction	14.9%	13.9%	6.7%
Miscellaneous	48.9%	48.0%	46.9%
Recommended use	30.9%	28.2%	44.1%
Value	91.5%	87.6%	84.2%

Value of the Reviewed Products

Themes on value included narrative segments about each product's efficacy, price, and worth, as well as the user's willingness to use the product again. Based on the 40 longest reviews for each product (n=120), regardless of the product, most reviewers were hopeful and elated to finally feel relief from BV. For example, a VC reviewer said:

> *I am so happy…I hope this is the end to my BV nightmare.*

Another said:

> *This is the best I have felt in 6 months.*

Similarly, an OM reviewer said:

> *I am so happy with the results of this medicine…I just hope I never get BV again!!!!!*

A few reviewers with recurrent BV tried to moderate expectations. For example, one said:

> *I like [VC] better than the alternatives. Do not expect anything to be a miracle cure for this problem.*

Another reviewer said that after failing multiple rounds of an oral treatment she "was told to do [VM] twice a week for 6 months," which "kept it at bay but [I] never got rid of it."

VC and VM reviewers reported symptom relief within 1-2 days compared to 2-3 days for OM reviewers. Positive value statements for VC included testimonies such as:

> *[The] next day, I felt so much better…I was dancing around the house…with not a symptom of any kind.*
>
> *[I] feel so clean and better now than ever before without any side effects.*

More than half of the VC users said they would take the medication again, whereas at least 4 VC reviewers said they would not recommend it because it resulted in a severe and treatment-resistant yeast infection. Only 1 VC reviewer stopped using it because of "'wicked' stomach pain and diarrhea" as well as "feeling like my vajayjay was on FIRE…" within

XSL•FO
RenderX

minutes of inserting the VC ovule. Three VC reviewers complained that the product was expensive, even with insurance or a coupon.

Of the 40 reviews analyzed, 14 (35%) VM reviewers reported symptom relief, 15 (37.5%) said they would use VM again, 6 (15%) said they would try something else, 2 (5%) found no relief, and 2 (5%) felt it was too soon to say whether the treatment worked for them. In addition, 1 (2.5%) VM reviewer gave it a 5-star rating "simply because it worked exactly as intended and cleared up BV probably within the first two days…everything is brand new and healthy," while 6 (15%) VM reviewers said it caused uncomfortable side effects. Those with burning and yeast infections said they would not use the medication again. Only 1 (2.5%) reviewer said she stopped using the product due to side effects, such as cramps and anxiety, but noted she "made the mistake of reading reviews" before starting treatment and was "terrified" to take VM. Three reviewers finished the treatment, however, but stated they would not take it again. One specifically disagreed with taking a vaginal medication:

> *I do not ever want to take this med again, and if I have to, I will insist on oral form. Why would they even make a med that had to be inserted vaginally when that area is so irritated from the infection?*

Although 10 OM reviewers reported symptom relief, the vast majority of reviews contained negative value statements about its side effects—described by one as "hell on earth." More OM reviewers said they would not use it again due to side effects compared to the other products. A few OM reviewers said the treatment did not work, as their symptoms persisted or worsened. Some reviewers commented that OM is expensive and not worth it, given its poor effectiveness and side effects. One said:

> *My advice…pay more out of pocket and opt for the vaginal gel as it was more effective for me and few to none with regards to side effects.*

Another user warned:

> *If you have BV, please ask for the gel. It is expensive, but this pill is so hard on your body.*

Three OM reviewers discussed stopping the medication. One reviewer skipped multiple doses because she was "scared to take [OM] again" due to vaginal swelling and irritation. The remaining 2 reviewers considered stopping OM. One had an adverse reaction to alcohol:

> *I am tempted to stop at 7 [of 10 days of treatment], but I really want this infection gone at once.*

Another reviewer said:

> *I only have 5 days left, but really debating on whether or not to go ask for something different…[because the side effects were] making me miserable!!*

Side Effects of the Reviewed Products

Among all three products, a portion of reviewers reported side effects. Specifically, 33 reviewers mentioned experiencing side effects while taking OM (vs 2 who reported none, 5 not mentioned), 27 for VM (vs 1 none, 12 not mentioned), and 25 for VC (vs 10 none, 5 not mentioned). OM reviewers described a wide variety of severe and plentiful side effects. Most OM reviewers experienced multiple side effects from flulike symptoms, gastrointestinal (GI) issues, nausea and vomiting, chest discomfort, mental health issues (depression, anxiety), bleeding, pain, hair loss, fatigue, brain fog, and, to a lesser extent, discharge, yeast infections, and changes in urine. Avoiding alcohol "was the hardest part" for many reviewers, especially daily drinkers who "love a glass of wine with dinner every night."

Similar to OM, a majority of VM users also experienced side effects. Common VM side effects included discharge described as white, clumpy, or stringy, and rarely bloody, followed by cramps and burning, which ranged from mild to extreme in intensity. A minority of VC reviewers experienced flulike symptoms, mental health issues (anxiety, depression), brain fog, nausea, changes in urine, GI discomfort, and headaches; only 1 reviewer described an allergic reaction (VM and OM). Side effects from VM and alcohol use were mixed, with 1 reviewer saying they drank alcohol and did not experience vomiting. Only 2 VM reviewers experienced a severe yeast infection.

The most common side effects reported by VC reviewers were leakage/discharge, followed by yeast infection, burning/irritation, GI discomfort, itching, headaches, and fatigue. Severe and treatment-resistant yeast infections were most common in VC reviewers and the reason some VC reviewers "regret not choosing" a different BV product, such as metronidazole. Additionally, many of the reviewers who only experienced leakage/discharge felt this was typical of vaginal medications and were not bothered by this side effect, especially when their symptoms resolved.

Characteristics of the Reviewed Products

Some reviews discussed characteristics of the products (attributes, taste, consistency, method of delivery, packaging, etc) that reviewers liked and did not like. An overwhelming number of OM reviewers complained about the product's "horrible," "nasty," and "terrible" taste. Some OM reviewers reported that it dissolved so quickly in their mouth that they could not swallow it fast enough to avoid the taste. Overall, reviewers felt the vaginal products (ie, VC and VM) were easy to use. A few VM reviewers likened it to "a thin tampon," and one said:

> *If you are used to using tampons, this will be very easy to apply.*

Not all reviewers preferred vaginal medications. One VM reviewer who does not use tampons said she was "not too fond of inserting things inside of my vagina…" but that "the gel is so smooth it almost feels like nothing." A few VC and VM reviewers felt the products have a pleasant cooling or smooth sensation. One VC reviewer said:

> *The only annoyance I had was that the suppository doesn't stay in the vagina, and I had to push it back in and lie down until it dissolved.*

Another VM reviewer said she "…could have done without the mess of the gel and the applicators, though"; however, 2 VM reviewers remarked that the "gel did not come out" after

XSL•FO
RenderX

insertion, and 1 said she "even had sex one night after using and neither of us noticed the gel." One VC reviewer liked the one-dose treatment.

Cross-Product Discussions

A portion of VC, OM, and VM reviewers discussed using other products to treat BV. The largest group were reviewers who said they tried but did not find relief from OTC products—such as boric acid and probiotics—and sought a prescription medication. Many tried multiple prescription products. Another group of reviewers discussed taking an OTC product before, during, and after taking a prescription medication as a preventative. Finally, a minority of reviewers said they stopped taking a prescription product and only took OTC products. One VC reviewer said boric acid vaginal suppositories and once-daily acidophilus were "the best cure" because "it helps with pH and kills the odor." Other OM users also tried tea tree oil, vitamin C ovules, and vaginal washes, which did not alleviate their infection. Several reviewers said their doctors recommended eating yogurt or taking OTC products (probiotics, boric acid), in addition to prescription treatment. It was more common that reviewers would switch to VM after OM due to its "horrible, horrible side effects like nausea, bad taste in mouth, and a yeast infection." Another OM reviewer said:

> *If this is for BV, I urge you to try boric acid, probiotic, and the gel form rather than tablets. Alternatively, clindamycin can be considered as well.*

Among VC reviewers who mentioned cross products, the majority preferred VC over metronidazole because it has fewer side effects; one reviewer said, "Please give it a try if you are fed up with metronidazole. This is a great alternative."

Seeking Information and Guidance

Many reviewers said they read "all the reviews" before taking a product, which helped them decide whether to take it. Reading reviews left most reviewers "scared" or "terrified" to take the product, but only a few said they refused to take a product because of what they read. Others said they were "really glad I did not read the reviews" before using it. Some reported delaying treatment until the infection worsened. When they finally took it, many said they were relieved when the side effects were mild and worried over nothing.

Reviewers also learned about side effects and how to minimize them from reading reviews. One reviewer said:

> *I wish my [doctor] would have warned that a yeast infection was possible.*

Most OM reviews (n=24, 60%) focused on ways to minimize the "horrible"-tasting product.

Some used their reviews to ask questions. The most common question was how long after treatment should they abstain from having sex. A few reviewers had questions about whether their male partners should be tested or treated for BV:

> *I'm curious to know do my partner need to be tested before we have intercourse again? That wasn't explained to me at my doctor visit.*

Other reviewers turned to the reviews for guidance on treatment length, for example:

> *I was informed to complete a 10 day which is confusing because other reviews only state a 5-day term. Has anyone else been told to complete a 10 day?"*

Several wondered how long they would experience side effects, such as a metallic taste or leakage/discharge. For example, 1 VC user said:

> *It's a pasty stuff coming out...Is it the cream still coming out? How long does it take for it to totally come out?*

Airing Frustrations

State of the Science

Some reviewers expressed frustration about the state of BV treatments. One reviewer said:

> *I do not think the scientists work hard enough on a permanent BV solution. But at least, there are temporary fixes for now, I suppose.*

Another reviewer said:

> *How could something that seems to be so simple of an issue that so many women experience, be treated with medication that causes so many horrible side effects!?*

A third reviewer expressed a similar sentiment:

> *I think it is ridiculous that something that is so common in women is so hard to get rid of and this medication along with another one is the only thing that doctors can treat you with.*

Doctor Interactions

Multiple reviewers disclosed interactions they had with their doctors. Some reviewers reported their doctors refusing to test or treat them for BV. One reviewer said:

> *When I got tested from my gynecologist, she told me that I was borderline BV...so they refused to give me anything.*

Another reviewer said she was "switching to another OB gynecologist" because she was prescribed a product without being tested beforehand and suffered stomachaches, nausea, vomiting, and lethargy.

Others reported doctors encouraging them to complete treatment despite severe side effects. One reviewer appeared to have an allergic reaction to a product:

> *I went to my doctor because I was out of work because of this stuff. She had me finish the medicine.*

Some reviewers complained that their doctor or pharmacist did not but should have warned them to avoid alcohol and alcohol-containing beauty products (eg, sprays, lotions) and reported suffering extreme nausea and vomiting. Another reviewer who stopped taking a product due to side effects said her "VERY annoyed" doctor did not believe it could cause

XSL•FO
RenderX

diarrhea and "found it 'hard to believe' that the vaginal burning could be 'that bad'." When this reviewer asked for an alternative medication, her doctor "snapped that there 'aren't any' and told me to call my gynecologist."

Some reviewers were annoyed that they were not offered alternative treatments with fewer side effects sooner. For example, one reviewer who was taking a "disgusting tast[ing]" product that made her feel "horrible" for 6 months reported she "finally went to a gynecologist that listened…he prescribed me [VM] instead of pill form." Other reviewers reported having prolonged, repeat infections before their doctor switched medications. For example, one user said:

> *Apparently my body has become immune to [metronidazole], so my doctor wanted me to try* [clindamycin]

Others felt their doctor did not listen to their product preferences. For example, one reviewer said she asked her doctor "why he would not give me pills instead as I really did not want to insert anything into such an irritated area, but he insisted this was the best method."

Discussion

Principal Findings

This mixed methods study examined patient-reported sentiment toward three of the most prescribed BV medications using data from online drug review forums. Overall, findings show that people want a BV treatment that is easy to use, quickly alleviates symptoms, and has minimal side effects. Although current BV treatment prescribing patterns [33] recommend OM as a first-line therapy, followed by VM and VC, this study found a preference for vaginal products overall. Additionally, VC had higher star ratings compared to VM and OM, which was bolstered by qualitative findings suggesting that VM and VC are effective and typically relieve symptoms 1-2 days faster than OM. Moreover, the side effects of VM and VC are usually local (eg, leakage/discharge, yeast infection, and vaginal irritation), whereas OM is associated with a broader range of systemic and more severe side effects, including adverse alcohol interactions well documented in the literature [34].

Honoring patient preferences—including the desired route of administration and concerns about side effects—is essential for improving BV treatment satisfaction and outcomes [35,36]. Health care providers should engage patients in discussions about lifestyle factors that may influence treatment choice, such as alcohol use [34], and BV triggers [37], such as sexual and hygiene practices [3,4] and the risk of reinfection from an untreated partner [38,39].

Findings from this study, along with previous research, show that patients will use internet searches to self-diagnose BV [17] and seek self-help remedies, such as OTC medications [40], even when this leads to misinformation [41]. This is unsurprising, given the negative psychosocial [42] and quality-of-life effects associated with BV [18,19]. For example, studies have shown that as many as 50% of BV patients mistakenly believe they have a yeast infection and take OTC antifungal treatments [43]. The remaining 50%, often those with recurrent BV who correctly identify their condition, frequently try OTC treatments, probiotics, or home remedies before seeking professional care [44].

Health care providers have a unique opportunity to educate patients about BV, dispel common myths, and inform them about available treatments. For instance, some OM reviewers expressed surprise and satisfaction upon learning that metronidazole is available in vaginal form. Others wanted their providers to proactively discuss side effects, particularly avoiding alcohol and alcohol-containing beauty products while on metronidazole. Patients also need information about when they resume sexual activity. Comprehensive education, whether provided in clinical encounters or through marketing and outreach efforts, is needed to help patients make informed decisions without reliance on online forums, where misinformation can be found [41]. At a broader level, public health campaigns can help reduce the stigma surrounding BV, overcome feelings of embarrassment associated with the disease, and empower patients with accurate information and support.

Strengths and Limitations

Findings should be interpreted while considering the study's strengths and limitations. Although the study relied on cross-sectional data, potentially from patients with a greater willingness to share their experiences via online platforms, it is a strength that the study used a large sample and the proportion of products reviewed largely mimics guideline-prescribing patterns. Another limitation is that the data are self-reported and assumed to be accurate, truthful, and from the person who used the product as directed, which cannot be verified. It is not known whether reviewers were paid money or given free products for their review. To submit a review, users must create an account, which is a strength as this may deter bots or fake reviews. Furthermore, each website uses its own security measures to limit fake/bot-generated reviews.

Another limitation of the study is that the NRC lexicon is not specific to BV, medical conditions, or pharmaceutical treatments. Therefore, BV-specific terms may not be captured in these analyses. Given the exploratory nature of this study, future studies should consider using other internet sources where patients provide robust information about their health care needs, symptoms, experiences, and recommendations. Combining data from several websites could be used to validate or create a lexicon that is more specific to a given medical condition.

A strength of the study is that it leveraged both qualitative and quantitative data to include patient-reported ratings and SA scores. Including a traditional qualitative analysis may have helped overcome critiques about the accuracy of SA approaches, which depend on the lexicon used, how well it matches subjective interpretation, and complex speech patterns (sarcasm, irony, negation, misspellings, etc). At multiple points, we used a human assessment of reviews to increase rigor and validity. As is the case with all qualitative studies, the sample is not representative of the population, and we only analyzed a subset of the 40 longest reviews per product. Although only one author (AC) performed the qualitative analysis, a strength is that the themes aligned with the double-coded and verified classifications. Because a reviewer could write anything they

XSL•FO
RenderX

wanted, caution should be used when interpreting counts in the qualitative results, as some only discussed their experience with BV, not information about the reviewed product. Although we selected the 40 longest reviews for each product for the qualitative analysis to focus on the richest data, these reviews may have been different (eg, represent more extreme experiences) than an average review. Due to the study design, we were unable to engage with the review to ask follow-up or clarifying questions.

Conclusion

Despite being the most common cause of vaginal discharge in people of childbearing age in the United States, innovation toward treatment options for BV are limited. This study's findings substantiate the need for a BV treatment that that is easy to use, quickly alleviates symptoms, and has minimal side effects. Considering how online resources are a go-to source for patient health information, these findings also underscore the need for comprehensive education to help patients differentiate facts from myths related to BV disease. Finally, considering patient preferences when prescribing BV treatment options is important for achieving treatment success and ensuring satisfaction.

Acknowledgments

The authors acknowledge and thank Sandra Wangeshi for her contributions to data acquisition and coding. In addition, editorial support was provided by Dr Rajiv Ahlawat and Dr Tanya Debnath (Tata Consultancy Services, India), funded by Organon & Co.

Authors' Contributions

EW, AC, CC, JR, and JA were responsible for study conception and design, data interpretation, critically revising the manuscript, and providing approval for the final version of the manuscript. EW, AC, and CC were responsible for data acquisition and analysis. EW and AC were responsible for drafting the manuscript.

Conflicts of Interest

EW and JR are employees of Organon & Co and may hold stock or shares. AC, CC, JA, and SW are external consultants for Organon & Co, which provided consulting fees in connection with this study and the development of this manuscript. AC is the owner of Rogue Scholar Consulting, which received financial support from Organon & Co in connection with this study and the development of this manuscript. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Multimedia Appendix 1
Positive emotion word proportions by product.
[PNG File , 282 KB - ojphi_v17i1e71720_app1.png]

Multimedia Appendix 2
Negative emotion word proportions by product.
[PNG File , 260 KB - ojphi_v17i1e71720_app2.png]

References

1. Morris M, Nicoll A, Simms I, Wilson J, Catchpole M. Bacterial vaginosis: a public health review. BJOG 2001 May;108(5):439-450 [FREE Full text] [doi: 10.1111/j.1471-0528.2001.00124.x] [Medline: 11368127]
2. Department of Health and Human Services. Bacterial vaginosis: a fact sheet from the Office on Women's Health. Office on Women's Health. 2015. URL: https://owh-wh-d9-dev.s3.amazonaws.com/s3fs-public/documents/fact-sheet-bacterial-vaginosis.pdf [accessed 2025-09-17]
3. Ravel J, Moreno I, Simón C. Bacterial vaginosis and its association with infertility, endometritis, and pelvic inflammatory disease. Am J Obstet Gynecol 2021 Mar;224(3):251-257 [FREE Full text] [doi: 10.1016/j.ajog.2020.10.019] [Medline: 33091407]
4. Fethers K, Fairley C, Hocking J, Gurrin L, Bradshaw C. Sexual risk factors and bacterial vaginosis: a systematic review and meta-analysis. Clin Infect Dis 2008 Dec 01;47(11):1426-1435. [doi: 10.1086/592974] [Medline: 18947329]
5. Brusselmans J, De Sutter A, Devleesschauwer B, Verstraelen H, Cools P. Scoping review of the association between bacterial vaginosis and emotional, sexual and social health. BMC Womens Health 2023 Apr 07;23(1):168 [FREE Full text] [doi: 10.1186/s12905-023-02260-z] [Medline: 37029382]
6. Peipert J, Lapane K, Allsworth J, Redding C, Blume J, Stein M. Bacterial vaginosis, race, and sexually transmitted infections: does race modify the association? Sex Transm Dis 2008 Apr;35(4):363-367. [doi: 10.1097/OLQ.0b013e31815e4179] [Medline: 18360319]

XSL•FO
RenderX

7. Atashili J, Poole C, Ndumbe P, Adimora A, Smith J. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. AIDS 2008 Jul 31;22(12):1493-1501 [FREE Full text] [doi: 10.1097/QAD.0b013e3283021a37] [Medline: 18614873]
8. Koumans E, Sternberg M, Bruce C, McQuillan G, Kendrick J, Sutton M, et al. The prevalence of bacterial vaginosis in the United States, 2001-2004; associations with symptoms, sexual behaviors, and reproductive health. Sex Transm Dis 2007 Nov;34(11):864-869 [FREE Full text] [doi: 10.1097/OLQ.0b013e318074e565] [Medline: 17621244]
9. Knochenhauer H, Lim S, Havrilesky L, Dotters-Katz S. The cost-effectiveness of bacterial vaginosis testing prior to delivery. Am J Obstet Gynecol 2022 Feb;226(2):297 [FREE Full text] [doi: 10.1016/j.ajog.2021.11.1289]
10. Powell A, Ghanem KG, Rogers L, Zinalabedini A, Brotman RM, Zenilman J, et al. Clinicians' Use of Intravaginal Boric Acid Maintenance Therapy for Recurrent Vulvovaginal Candidiasis and Bacterial Vaginosis. Sex Transm Dis 2019 Dec;46(12):810-812 [FREE Full text] [doi: 10.1097/OLQ.0000000000001063] [Medline: 31663976]
11. Reiter S, Kellogg Spadt S. Bacterial vaginosis: a primer for clinicians. Postgrad Med 2019 Jan 30;131(1):8-18 [FREE Full text] [doi: 10.1080/00325481.2019.1546534] [Medline: 30424704]
12. Peebles K, Velloza J, Balkus JE, McClelland RS, Barnabas RV. High Global Burden and Costs of Bacterial Vaginosis: A Systematic Review and Meta-Analysis. Sex Transm Dis 2019 May;46(5):304-311 [FREE Full text] [doi: 10.1097/OLQ.0000000000000972] [Medline: 30624309]
13. Watkins E, Chow CM, Lingohr-Smith M, Lin J, Yong C, Tangirala K, et al. Treatment patterns and economic burden of bacterial vaginosis among commercially insured women in the USA. J Comp Eff Res 2024 Jan;13(1):e230079 [FREE Full text] [doi: 10.57264/cer-2023-0079] [Medline: 38099520]
14. Bradshaw CS, Morton AN, Hocking J, Garland SM, Morris MB, Moss LM, et al. High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. J Infect Dis 2006 Jun 01;193(11):1478-1486 [FREE Full text] [doi: 10.1086/503780] [Medline: 16652274]
15. Sobel JD, Schmitt C, Meriwether C. Long-term follow-up of patients with bacterial vaginosis treated with oral metronidazole and topical clindamycin. J Infect Dis 1993 Mar 01;167(3):783-784 [FREE Full text] [doi: 10.1093/infdis/167.3.783] [Medline: 8440952]
16. Surapaneni S, Akins R, Sobel JD. Recurrent bacterial vaginosis: an unmet therapeutic challenge. Experience with a combination pharmacotherapy long-term suppressive regimen. Sexual Trans Dis 2021 Mar 22;48(10):761-765 [FREE Full text] [doi: 10.1097/olq.0000000000001420] [Medline: 34110746]
17. Velmahos AH, Cooley Demidkina B, Mitchell CM. Comparison of bacterial vaginosis symptoms reported in social media vs. those reported by patients. Front Reprod Health 2025 Feb 28;7:1549331 [FREE Full text] [doi: 10.3389/frph.2025.1549331] [Medline: 40093028]
18. Chow K, Wooten D, Annepally S, Burke L, Edi R, Morris SR. Impact of (recurrent) bacterial vaginosis on quality of life and the need for accessible alternative treatments. BMC Womens Health 2023 Mar 18;23(1):112 [FREE Full text] [doi: 10.1186/s12905-023-02236-z] [Medline: 36934289]
19. Payne SC, Cromer PR, Stanek MK, Palmer AA. Evidence of African-American women's frustrations with chronic recurrent bacterial vaginosis. J Am Acad Nurse Pract 2010 Feb 28;22(2):101-108 [FREE Full text] [doi: 10.1111/j.1745-7599.2009.00474.x] [Medline: 20132368]
20. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Patient-focused drug development: collecting comprehensive and representative input. Guidance for industry, food and drug administration staff, and other stakeholders. Food and Drug Administration. 2020 Jun. URL: https://www.fda.gov/media/139088/download [accessed 2025-09-17]
21. Mao Y, Liu Q, Zhang Y. Sentiment analysis methods, applications, and challenges: a systematic literature review. J King Saud Univ Comput Inf Sci 2024 Apr;36(4):102048 [FREE Full text] [doi: 10.1016/j.jksuci.2024.102048]
22. Mohammad S, Turney P. Emotions evoked by common words and phrases: using Mechanical Turk to create an emotion lexicon. 2010 Presented at: NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text; June 2010; Los Angeles, CA URL: https://aclanthology.org/W10-0204/ [doi: 10.4324/9780429508059-1]
23. Jones P. Design for Care: Innovating Healthcare Experience. New York: Rosenfeld Media; 2013.
24. Accurate, independent and on demand - Drugs.com. IBM Corporation. 2021 Oct. URL: https://www.ibm.com/case-studies/drugsdotcom [accessed 2025-07-15]
25. Plutchik R. The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. Am Sci 2001;89(4):344-350 [FREE Full text] [doi: 10.1511/2001.28.739]
26. Semeraro A, Vilella S, Ruffo G. PyPlutchik: visualising and comparing emotion-annotated corpora. PLoS One 2021 Sep 1;16(9):e0256503 [FREE Full text] [doi: 10.1371/journal.pone.0256503] [Medline: 34469455]
27. Glaser BG, Strauss AL, Strutzel E. The discovery of grounded theory; strategies for qualtiative research. Nurs Res 1968;17(4):364 [FREE Full text] [doi: 10.1097/00006199-196807000-00014]
28. Kim H, Sefcik JS, Bradway C. Characteristics of qualitative descriptive studies: a systematic review. Res Nurs Health 2017 Feb 30;40(1):23-42 [FREE Full text] [doi: 10.1002/nur.21768] [Medline: 27686751]
29. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov 27;26(13):1753-1760 [FREE Full text] [doi: 10.1177/1049732315617444] [Medline: 26613970]

XSL•FO
RenderX

30. LaDonna KA, Artino AR, Balmer DF. Beyond the guise of saturation: rigor and qualitative interview data. J Grad Med Educ 2021 Oct;13(5):607-611 [FREE Full text] [doi: 10.4300/JGME-D-21-00752.1] [Medline: 34721785]
31. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340 [FREE Full text] [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
32. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
33. Bacterial vaginosis - STI treatment guidelines. Division of STD Prevention, Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/std/treatment-guidelines/bv.htm [accessed 2025-07-15]
34. Mergenhagen KA, Wattengel BA, Skelly MK, Clark CM, Russo TA. Fact versus fiction: a review of the evidence behind alcohol and antibiotic interactions. Antimicrob Agents Chemother 2020 Feb 21;64(3):e02167-19 [FREE Full text] [doi: 10.1128/aac.02167-19] [Medline: 31871085]
35. Muzny CA, Kardas P. A narrative review of current challenges in the diagnosis and management of bacterial vaginosis. Sexual Trans Dis 2020 Mar 27;47(7):441-446 [FREE Full text] [doi: 10.1097/olq.0000000000001178] [Medline: 32235174]
36. Chavoustie SE, Eder SE, Koltun WD, Lemon TR, Mitchell C, Nyirjesy P, et al. Experts explore the state of bacterial vaginosis and the unmet needs facing women and providers. Int J Gynaecol Obstet 2017 May 06;137(2):107-109 [FREE Full text] [doi: 10.1002/ijgo.12114] [Medline: 28262922]
37. Bilardi J, Walker S, Mooney-Somers J, Temple-Smith M, McNair R, Bellhouse C, et al. Women's views and experiences of the triggers for onset of bacterial vaginosis and exacerbating factors associated with recurrence. PLoS One 2016 Mar 1;11(3):e0150272 [FREE Full text] [doi: 10.1371/journal.pone.0150272] [Medline: 26930414]
38. Vodstrcil LA, Muzny CA, Plummer EL, Sobel JD, Bradshaw CS. Bacterial vaginosis: drivers of recurrence and challenges and opportunities in partner treatment. BMC Med 2021 Sep 02;19(1):194 [FREE Full text] [doi: 10.1186/s12916-021-02077-3] [Medline: 34470644]
39. Bilardi JE, Walker SM, Temple-Smith MJ, McNair RP, Mooney-Somers J, Vodstrcil LA, et al. Women view key sexual behaviours as the trigger for the onset and recurrence of bacterial vaginosis. PLoS One 2017 Mar 9;12(3):e0173637 [FREE Full text] [doi: 10.1371/journal.pone.0173637] [Medline: 28278277]
40. Bilardi J, Walker S, McNair R, Mooney-Somers J, Temple-Smith M, Bellhouse C, et al. Women's management of recurrent bacterial vaginosis and experiences of clinical care: a qualitative study. PLoS One 2016 Mar 24;11(3):e0151794 [FREE Full text] [doi: 10.1371/journal.pone.0151794] [Medline: 27010725]
41. Borges do Nascimento IJ, Beatriz Pizarro A, Almeida J, Azzopardi-Muscat N, André Gonçalves M, Björklund M, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022 Sep 01;100(9):544-561 [FREE Full text] [doi: 10.2471/blt.21.287654] [Medline: 36062247]
42. Irving G, Miller D, Robinson A, Reynolds S, Copas AJ. Psychological factors associated with recurrent vaginal candidiasis: a preliminary study. Sex Transm Infect 1998 Oct 01;74(5):334-338 [FREE Full text] [doi: 10.1136/sti.74.5.334] [Medline: 10195028]
43. Ferris D, Nyirjesy P, Sobel JD, Soper D, Pavletic A, Litaker MS. Over-the-counter antifungal drug misuse associated with patient-diagnosed vulvovaginal candidiasis. Obstet Gynecol 2002 Mar;99(3):419-425 [FREE Full text] [doi: 10.1016/s0029-7844(01)01759-8] [Medline: 11864668]
44. Majerovich JA, Canty A, Miedema B. Chronic vulvar irritation: could toilet paper be the culprit? Can Fam Physician 2010 Apr;56(4):350-352 [FREE Full text] [Medline: 20393094]

Abbreviations

AE: adverse event
BV: bacterial vaginosis
FDA: Food and Drug Administration
GI: gastrointestinal
NRC: National Research Council Canada
OM: oral metronidazole
OTC: over the counter
SA: sentiment analysis
VC: vaginal clindamycin
VM: vaginal metronidazole

XSL•FO
RenderX

Edited by E Mensah; submitted 24.01.25; peer-reviewed by JL Karim, J De Souza, E Young, Y Jiang; comments to author 28.05.25; revised version received 11.08.25; accepted 04.09.25; published 10.10.25.

Please cite as:
Watkins E, Cimino AN, Culbertson C, Raymaker J, Amico JR
Understanding Patient Perceptions of Bacterial Vaginosis Treatments: Mixed Methods Sentiment Analysis Study of Online Drug Review Forums
Online J Public Health Inform 2025;17:e71720
URL: https://ojphi.jmir.org/2025/1/e71720
doi:10.2196/71720
PMID:41072007

©Eren Watkins, Andrea N Cimino, Curtis Culbertson, Juliana Raymaker, Jennifer R Amico. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 10.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Quantifying Patient Demand for Orthopedics Care by Region Through Google Trends Analysis: Descriptive Epidemiology Study

Abram Qiu*, BS; Kristopher Meadows*, BS; Fei Ye*, BS; Osasu Iyawe*, BBA; Kenneth Kenneth-Nwosa, MD

Department of Orthopaedics, University of Texas Health Science Center, San Antonio, TX, United States

*these authors contributed equally

Corresponding Author:
Abram Qiu, BS
Department of Orthopaedics, University of Texas Health Science Center, San Antonio, TX, United States

Abstract

Background: There is a growing gap between the supply of surgeons and the demand for orthopedic services in the United States.

Objective: We analyzed publicly available online data to assess the correlation between the supply of orthopedic surgeons and patient demand across the United States. The geographic trends of this gap were assessed by using the relative demand index (RDI) to guide precision public health interventions such as resource allocation, residency program expansion, and workforce planning to specific regions.

Methods: The data used were from the US Census Bureau, Association of American Medical Colleges (AAMC) through their 2024 Electronic Residency Application Service (ERAS) directory, AAMC State Physician Workforce Data Report, and Google Trends. We calculated the normalized relative search volume (RSV) and the RDI and compared them to the densities of orthopedic surgeons across the United States. We examined the disparities with the Spearman rank correlation coefficient.

Results: The supply of orthopedic surgeons varied greatly across the United States, with a significantly higher demand for them in southern states (P=.02). The orthopedic surgeon concentration, normalized to the highest density, was the highest in Alaska (n=100), the District of Columbia (n=96), and Wyoming (n=72); and the lowest in Texas (n=0), Arkansas (n=6), and Oklahoma (n=64). The highest RDI values were observed in Utah (n=97), Florida (n=88), and Texas (n=83), while the lowest were observed in Alaska (n=0), the District of Columbia (n=5), and New Hampshire (n=7). The 7 states of Alaska, Maine, South Dakota, Wyoming, Montana, Delaware, and Idaho lacked orthopedic surgery residencies. In 2023, New York (n=19), Michigan (n=17), Ohio (n=17), Pennsylvania (n=16), and California (n=16) had the most residency programs. Demand and supply, represented by the RDI and orthopedic surgeon concentration, respectively, were strongly correlated negatively (ρ=−0.791, P<.001). States that were in the top quartile of residency programs (≥4 residency programs) exhibited a high demand for orthopedic surgeons (ρ=.6035, P=.02).

Conclusions: This study showed that regional disparities in access to orthopedic care can be addressed by increasing orthopedic residencies. The study highlights the novel application of the RDI to mapping the regional need for orthopedics, and this map allows for better targeted resource allocation to expand orthopedic surgery training.

(Online J Public Health Inform 2025;17:e63560) doi:10.2196/63560

KEYWORDS

orthopedics; geographic factors; health care disparities; medical schools; internship and residency; epidemiology; public health informatics; physicians; assessment of health care needs; resource allocation

Introduction

Orthopedic surgeries are among the most common medical procedures in the United States [1]. According to the American Academy of Orthopaedic Surgeons (AAOS), the number of orthopedic surgeons has not kept up with the need for orthopedic procedures. An AAOS fact sheet from March 2023 found that the current number of orthopedic surgeons must either double their total joint arthroplasty caseload or grow the number of surgeons by 10% every 5 years to meet the increasing demand [2]. This poses considerable challenges to the current state of health care, affecting the accessibility of services and increasing surgical workload amid physician burnout. We explored the geographic patterns of the supply of orthopedic surgeons and access to orthopedic services. The relative demand index (RDI) obtained in this study, derived from Google Trends data, provides a novel way to quantify the demand for orthopedic services considering the supply of residency-trained specialists across the United States. Unlike traditional measures of health care demand and supply, such as hospital utilization rates, census data, or claims data, Google Trends offers real-time insights into public interest and potential demand before it materializes

XSL•FO
RenderX

as service utilization. The RDI leverages this data to provide a dynamic and location-specific perspective, highlighting regions where interest in orthopedic services may signal unmet health care needs. This approach also addresses the key limitations of traditional methods. Conventional metrics often rely on retrospective, static datasets that may not fully capture emerging trends in demand or geographic disparities. By contrast, the RDI uses Google Trends to reflect immediate, user-driven interest, enabling a more granular understanding of demand. Using the RDI, we highlighted underserved regions and offer recommendations for targeted public health interventions. Additionally, we aimed to examine the interplay between this trend of shortages and the orthopedic specialty interest as both a potential contributing factor to and a consequence of this phenomenon. This will yield better insight into these challenges and guide the use of resources for existing and new orthopedic surgery residencies to fill the gap between the need for orthopedic surgeons and the supply of their services.

Methods

Data Collection

We gathered data from publicly available sources. The state population totals from 2020 to 2022 were obtained from the United States Census Bureau [3]. The location of every accredited orthopedic surgery residency program in the United States was collected from the Association of American Medical Colleges (AAMC) through their 2024 Electronic Residency Application Service (ERAS) directory [4]. The data of the current physician supply, number of undergraduate medical education students, and number of graduate medical education (GME) residents and fellows in the United States were gathered from the most up-to-date 2021 AAMC State Physician Workforce Data Report [5]. Lastly, Google Trends was used to search for data of "Orthopedic surgeons" from 2004 to 2023 to determine the demand for orthopedic surgeons in the United States [6].

Ethical Considerations

This study was conducted retrospectively and did not involve the use of any patient health record data. Therefore, it did not require assessment or approval by an institutional ethics review board.

Analytical Metrics

To quantify the relative demand for orthopedic surgeons across the country, we first calculated the relative search volume (RSV) in each state by using the search volume data collected from Google. Google Trends reported the data on a state-by-state basis, so we divided the frequency of specific searches for "Orthopedic surgeons" in every state by the monthly total number of all other searches performed in the state and then obtained the average value across the 2004–2023 time span. This number was then normalized on a scale from 0 to 100 to obtain the RSV, a metric that allowed us to compare search volumes across states with different populations. The higher the normalized RSV, the higher the interest in the topic.

The RSV was then used to calculate the RDI, our quantitative metric for relative orthopedic surgeon demand. We divided the number of practicing orthopedic surgeons per state, obtained from the 2021 AAMC State Physician Workforce Data Report, by the United States census state population to calculate the approximate concentration of orthopedic surgeons per 10,000 individuals. The concentration of orthopedic surgeons per 10,000 individuals was normalized on a scale from 0 to 100, where 100 represented the greatest concentration. Normalization was performed through the minimum–maximum method. For each value, the difference between the minimum and maximum numbers in the dataset was divided by the range of the dataset. We were able to use this method to enhance readability and preserve relationships between the original values because no outliers were identified in our dataset based on interquartile ranges. All the calculations are shown below.

$$\text{Relative search volume} = \text{Frequency of searches per state} / \text{Total monthly searches per state}$$

$$\text{Orthopedic surgeon concentration} = (\text{Orthopedic surgeons per state} / \text{State population}) \times 10{,}000$$

$$\text{Relative demand index} = \text{Relative search volume} / \text{Orthopedic surgeon concentration}$$

The RSV was then divided by the concentration of orthopedic surgeons to obtain the RDI, also normalized on a scale from 0 to 100. The normalized RDI measures the relative demand for orthopedic surgeons by incorporating both their supply and demand; thus, a state with a normalized RDI of 100 has the highest relative demand and a state with a normalized RDI of 0 has the lowest demand. As a result, it also serves as a marker of supply; the states with a higher normalized RDI have a relative undersupply of orthopedic surgeons whereas the states with a lower normalized RDI have a relative oversupply.

For example, if State A has an RSV of 0.005 (0.5% of all searches) and a surgeon concentration of 1 surgeon per 10,000 people (0.0001), the RDI is 50. In contrast, if State B has the same RSV but a higher surgeon concentration of 1 surgeon per 5000 people (0.0002), the RDI is 25. Even with identical demand, the higher RDI for State A reflects a greater relative demand due to fewer surgeons available per person.

To highlight the geographical variation observed between the supply of and demand for orthopedic surgeons, we subtracted the concentration of orthopedic surgeons per 10,000 people from the normalized RDI. This yielded the difference in interest and concentration, a metric from 0 to 100 that describes the relationship between the supply of and the demand for orthopedic surgeons. The larger the difference, the greater the disparity between the search volume of orthopedic surgeons and the orthopedic surgeon concentration in the state, and conversely, there is less of a disparity when the difference is smaller.

Statistical Analysis

Correlation analysis was run using Spearman rank correlation coefficient testing, which measures the strength and direction of association between two ranked variables. The statistical significance threshold was set as $P<.05$. Statistical analysis was performed with R software version 4.3.3 (R Foundation for Statistical Computing).

XSL•FO
RenderX

Results

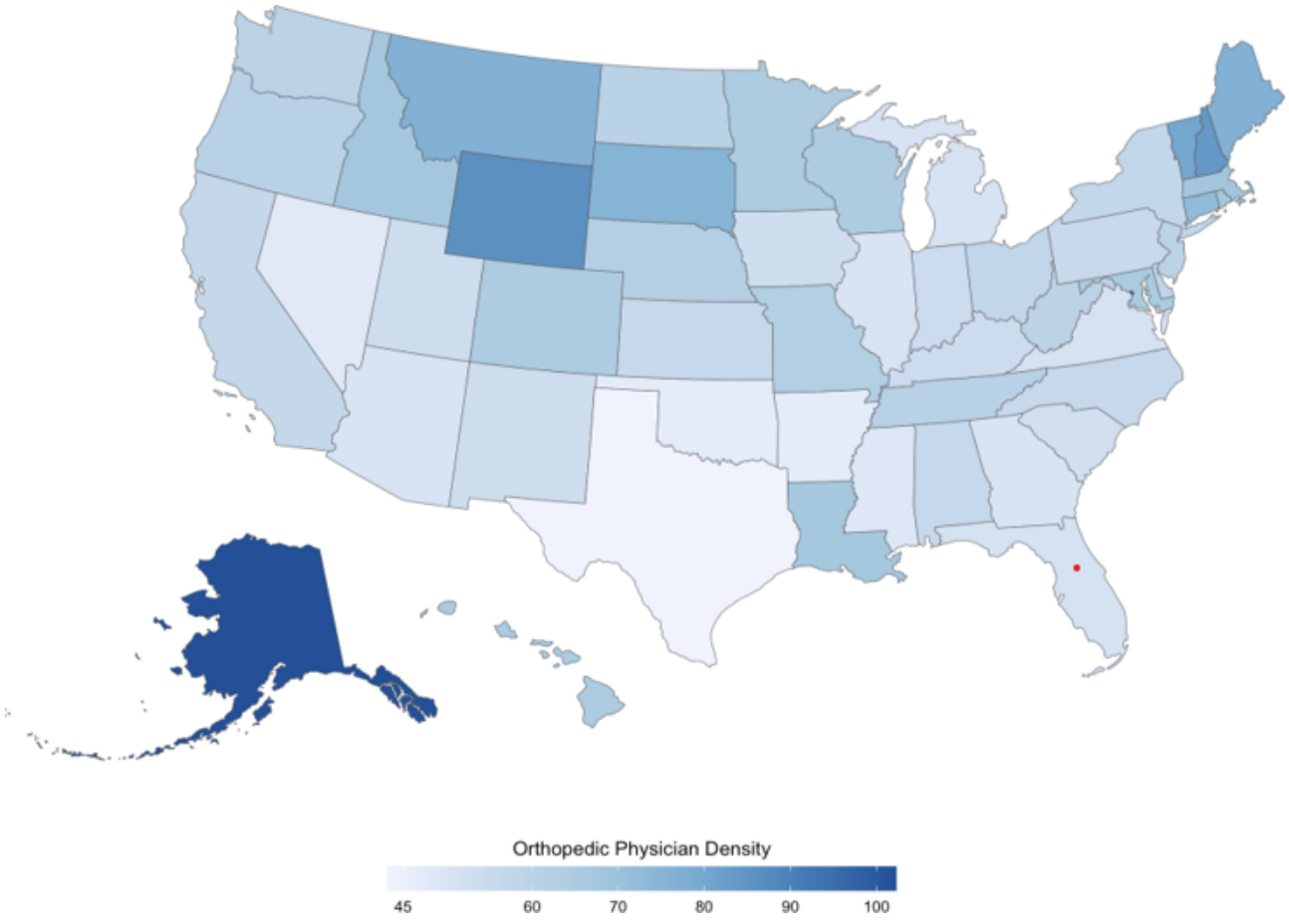
On preparing a graph of the normalized physician concentration versus the normalized RDI value across all states, we obtained a Spearman rank correlation coefficient of ρ=−0.791 (*P*<.001). This indicated that as the normalized RDI increased, the concentration of orthopedic surgeons decreased. On comparing the normalized RDI to the number of residency programs in the state, the coefficient was ρ=0.455 (*P*<.001). While it was a weaker association, it still indicated a high demand for orthopedic surgeons despite having many residency programs in the state. Alternatively, the large number of residency programs could be in response to the demand for orthopedics. Seven states had no residency programs for orthopedic surgery: Alaska, Maine, South Dakota, Wyoming, Montana, Delaware, and Idaho.

Our analysis of states below the 50th percentile of the RDI revealed an inverse relationship between the normalized RDI and physician concentration. This indicated that states with lower physician concentration exhibited a higher demand for orthopedic surgeons (ρ=−0.767, *P*<.001). Interestingly, within this subset of low-concentration states, a positive correlation was observed between the normalized RDI and the number of residency programs (ρ=0.510, *P*=.009). This suggested that increased residency training, or the number of residents who remain in the state they trained in, may not fully counteract the rising demand for orthopedic surgeons in these areas.

Furthermore, similar results were obtained when examining states with a normalized RDI exceeding the 50th percentile. The normalized RDI showed a negative correlation with physician concentration (ρ=−0.400, *P*=.044), again indicating a higher demand in areas with lower physician concentration. However, no statistically significant correlation was found between the normalized RDI and the number of residency programs in these high-demand states.

As shown in Figure 1, Alaska (n=100), the District of Columbia (n=96), and Wyoming (n=72) had the greatest orthopedic surgeon concentration; and states like Texas (n=0), Arkansas (n=6), and Oklahoma (n=64) had the lowest concentration. Notably, Alaska and Wyoming were also among the 7 states without orthopedic surgery residency programs. also demonstrates that areas in the Midwest and Northeast of the country had a relatively higher concentration of orthopedic surgeons.

Figure 1. Orthopedic physician surgeons per 10,000 individuals across the United States.

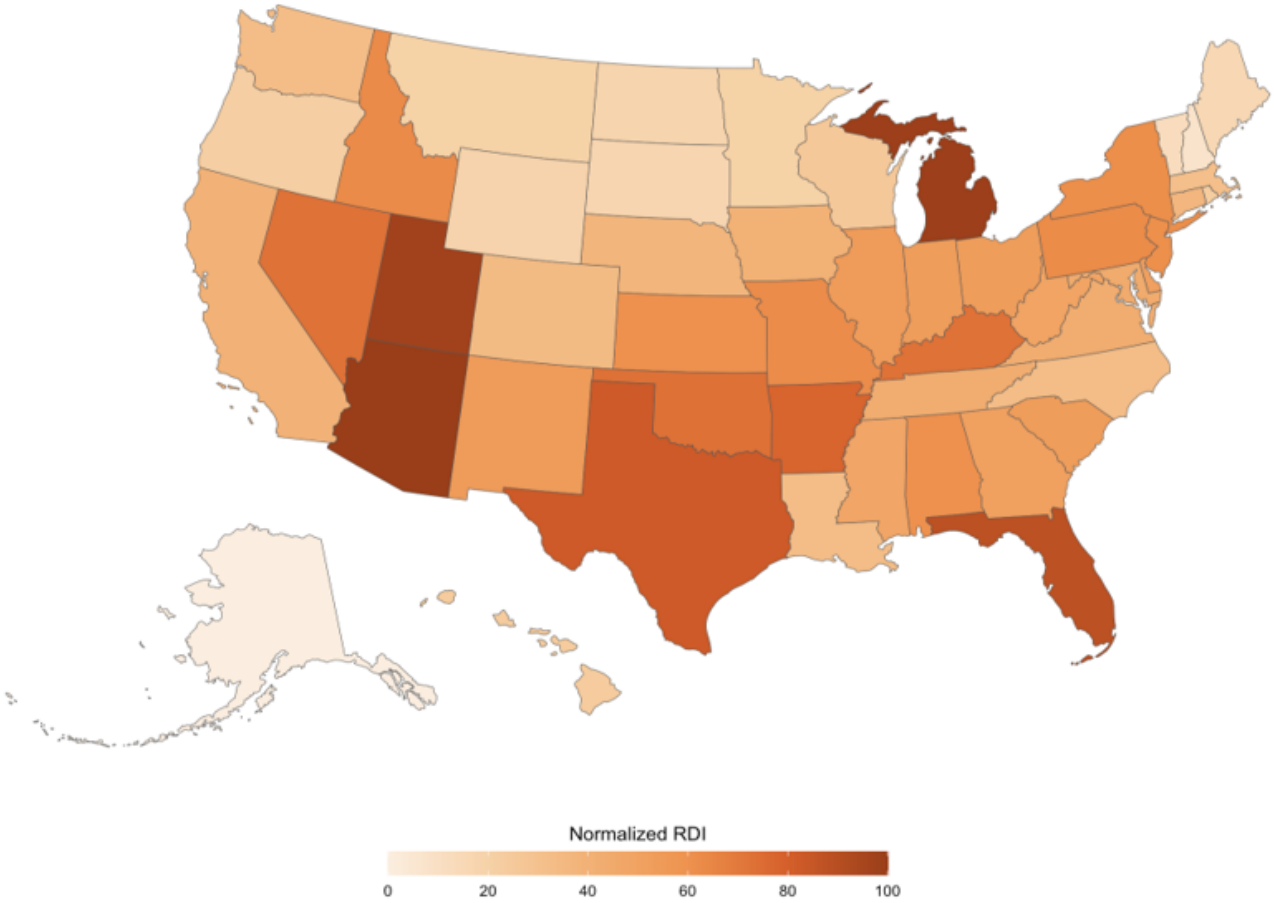


The states with the highest normalized RDI included Utah (n=97), Florida (n=88), and Texas (n=83), as highlighted in Figure 2. The territories with the lowest normalized RDI included Alaska (n=0), the District of Columbia (n=5), and New Hampshire (n=7). In contrast to the trend apparent with the orthopedic surgeon concentration, the states in the Midwest exhibited lower normalized RDI values, including states like South Dakota (n=0.17), North Dakota (n=18), Wyoming (n=19), Montana (n=20), and Minnesota (n=20). A few other states in

XSL•FO
RenderX

the Northeast also had low normalized RDI values, such as Vermont (n=13) and Maine (n=16).

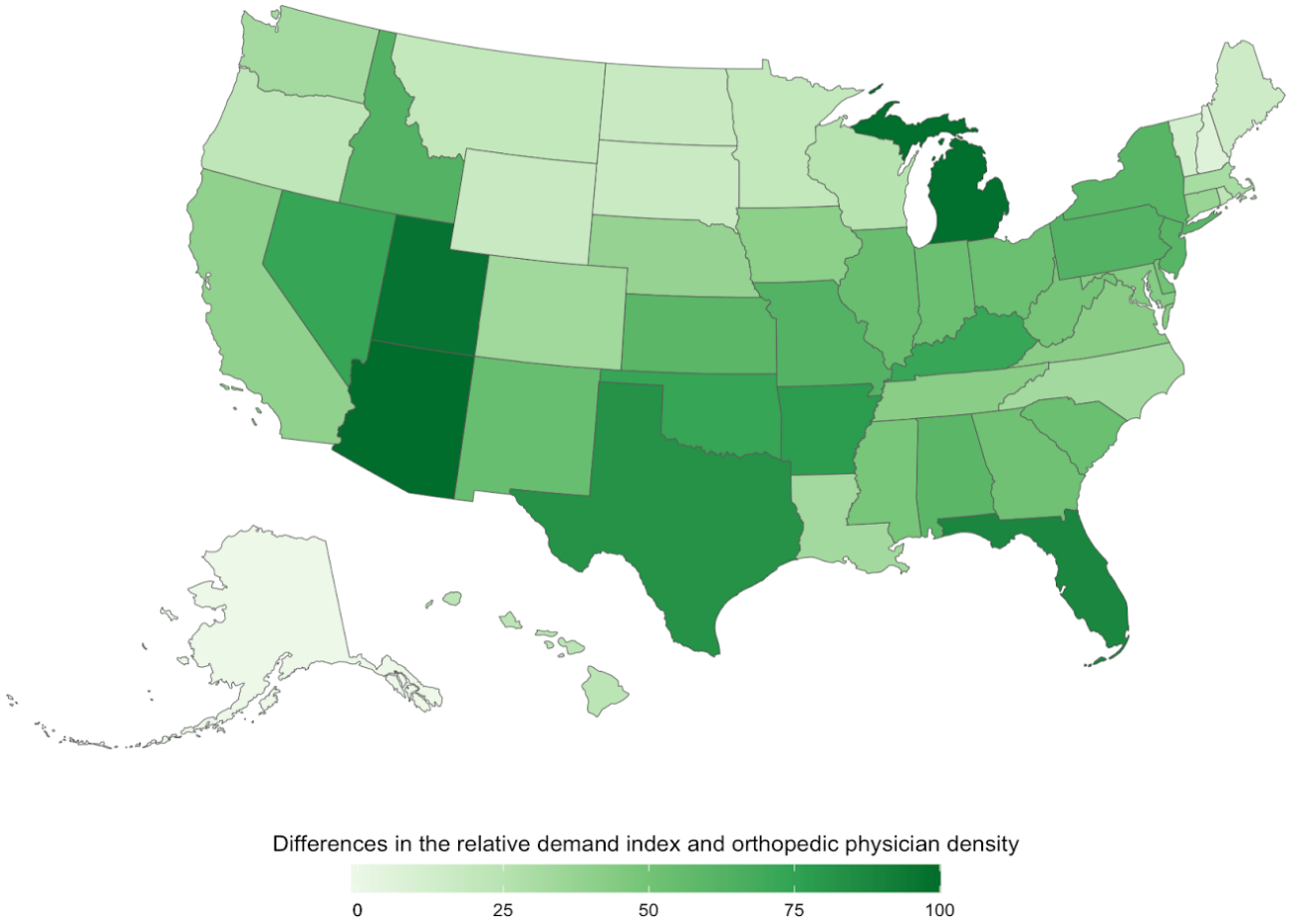
Figure 2. Normalized relative demand index (RDI) across the United States.



The differences in the RDI and orthopedic surgeon concentration are mapped in Figure 3. The states with the greatest difference between public interest and the supply of orthopedic surgeons were Arizona (n=87), Michigan (n=86), and Texas (n=83). The states with the lowest differences included Alaska (n=−100), the District of Columbia (n=−90), and New Hampshire (n=−62).

XSL•FO
RenderX

Figure 3. Differences in the relative demand index and orthopedic surgeon concentration across the United States.



Outside of Texas and Arizona, the country's Southwestern region had the greatest disparity between the RDI and orthopedic surgeon concentration. Meanwhile, much of the results in the Southeast appeared relatively uniform, indicating similar low differences in the relative demand interest and orthopedic surgeon concentration across the region.

As shown in Table 1, we categorized the states based on their normalized RDI and orthopedic surgeon concentration. The values ≥50 were categorized as "high" normalized RDI, while those <50 were designated as "low" normalized RDI. No states had both a high normalized RDI and a high orthopedic surgeon concentration.

XSL•FO
RenderX

Table . States categorized by combinations of high or low normalized relative demand index (RDI) and orthopedic surgeon concentration.

Variable	High normalized RDI	Low normalized RDI
High orthopedic surgeon concentration	• No states meet the criteria	• Alaska • District of Columbia • Maine • Montana • New Hampshire • South Dakota • Vermont • Wyoming
Low orthopedic surgeon concentration	• Alabama • Arizona • Arkansas • Delaware • Florida • Georgia • Idaho • Illinois • Indiana • Kansas • Kentucky • Michigan • Missouri • Nevada • New Jersey • New Mexico • New York • Ohio • Oklahoma • Pennsylvania • South Carolina • Texas • Utah	• California • Colorado • Connecticut • Hawaii • Iowa • Louisiana • Maryland • Massachusetts • Minnesota • Mississippi • Nebraska • North Carolina • North Dakota • Oregon • Rhode Island • Tennessee • Virginia • Washington • West Virginia • Wisconsin

The high normalized RDI and low orthopedic surgeon concentration category excluded states bordering the Pacific Ocean or the Canadian border (except for Michigan). Instead, this category of states was skewed towards the central and southern regions of the United States. Additionally, while some Eastern states (Delaware, New Jersey, and New York) were present, the northernmost portion of the Eastern seaboard (Maine, Vermont, and New Hampshire) lacked representation. The lower supply of orthopedic surgeons and greater demand for them demonstrate that this category of states would be ideal locations for new residencies. They may also be a good place for new orthopedic surgery graduates to find employment.

In the low normalized RDI and high orthopedic surgeon concentration category, all the states were in the northern half of the United States, concentrated in the Northeast (Maine, New Hampshire, and Vermont) and Northwest (Alaska, Montana, and Wyoming).

The low normalized RDI and low orthopedic surgeon concentration category of states was the most geographically diverse. This group included states from across the country, with representation from the West Coast (California, Oregon, and Washington), East Coast (Connecticut, Maryland, Massachusetts, Rhode Island, and Virginia), Midwest (Iowa, Minnesota, Nebraska, and Wisconsin), and South (Louisiana, Mississippi, North Carolina, and Tennessee).

The data sources and the ways in which the data were used are shown in Table 2.

XSL•FO
RenderX

Table . Sources of data collected and the ways they were integrated into our analysis.

Data sources	Use of data
AAMC ERAS[a] directory	Identified the number and location of all orthopedic surgery residency programs in the United States. Used the residency program data to compare to demand for orthopedic surgeons.
2021 AAMC State Physician Workforce Data	Provided the number of practicing orthopedic surgeons in the country. Used to quantify the concentration of orthopedic surgeons to create the relative demand index as well as the difference in interest and concentration of orthopedic surgeons.
United States Census Bureau	Provided state population totals during 2020 - 2022. Used to calculate concentration in the RSV[b] calculations.
Google Trends	Provided the volume of search data for "Orthopedic surgeons" during 2004 - 2023. Used to calculate the RSV.

[a]AAMC ERAS: Association of American Medical Colleges through the Electronic Residency Application Service.

[b]RSV: relative search volume.

Discussion

Principal Findings and Comparison With Previous Studies

In this study, the observed relationship between the number of orthopedic residency programs and the normalized RDI is of interest. Certain states exhibited a scarcity of orthopedic residency programs and a limited demand for orthopedic services, such as South Dakota, North Dakota, Wyoming, Montana, and Minnesota. Conversely, a discernible trend emerged wherein states with a greater demand for orthopedic services showed a proportionate increase in the concentration of residency programs. This trend is found in other specialties as well. For example, in psychiatry, Pheister et al [7] found that 84.2% of residency program directors who started or expanded programs listed the shortage of local psychiatrists as the main driver.

We found a high orthopedic surgeon concentration and relatively low normalized RDI in the Midwest. This lower normalized RDI may signify that the demand for orthopedic surgeons is met with a greater concentration of such surgeons. As a region with a smaller concentration of metropolitan areas, finding that this market was saturated was particularly intriguing due to the well-documented trend of physicians preferring to practice in more urban regions. For example, in a systematic literature review, Cyr et al [8] concluded that specialists gravitate toward urbanized areas with more potential patients to support their practice.

Our approach has been validated by Ikpeze et al [9], Blau et al [10], and Akosman et al [11]. Combining Google Trends data and the number of orthopedic surgeons and residency programs from the AAMC allowed us to interpret the geographical distribution of demand. Ikpeze et al [9] examined the relationship between the orthopedic surgery applicant pool in the United States and Internet queries for orthopedic residency. They found that searches for "orthopedic residency" and "orthopedic salary" increased dramatically from 2004 to 2015. The National Resident Matching Program database also showed a consistent yearly increase in orthopedic surgery applicants. This positive correlation between an official database and Google Trends demonstrates its validity in gauging public interest in a topic [9].

Google Trends analysis has explored other specialties and topics outside of orthopedic surgery. Blau et al [10] looked at all searches for "plastic surgery" over 12 months, and similar to our study, used RSV and surgeons-per-capita value, or "surgical concentration." Their Google RSV divided by this concentration yielded a "surgical demand index" for each state, which is similar to and validates our study's normalized RDI. Their "surgical demand index" was the greatest in Wyoming, Oklahoma, and Arkansas; and the smallest in Oregon, Virginia, and Connecticut [10].

Akosman et al [11] performed a similar analysis for the ophthalmology specialty; they suggested that states with a higher normalized RDI may have an unmet medical need for ophthalmic care. The highest normalized RDI was found in South Dakota, Delaware, Michigan, and Arizona. The lowest normalized RDI was in Washington DC, Hawaii, Oregon, and Montana. In this study for the orthopedic specialty, we found the lowest normalized RDI to be in the District of Columbia, Alaska, and New Hampshire.

Career factors that may have influenced the geographical distribution include the supply of advanced medical facilities, renowned orthopedic institutions, research opportunities, and access to innovative technology. Additionally, lifestyle factors, including the overall quality of life, salary, cost of living, and recreational amenities in a particular state, are influential variables as well. It is difficult to objectively explain the drive to or away from certain states. The insight our data provides regarding the demand for orthopedic services can guide new orthopedic surgeons to places in the United States where they are needed. In addition, it could also inform where to establish new residencies, as well as how existing residency programs should invest in recruiting medical students to their state. The limitations to establishing such programs include high costs and complex requirements by the Accreditation Council for Graduate Medical Education (ACGME), which can be overcome by acquiring increased GME funding and partnering with hospitals, universities, and policymakers to reduce financial burden and facilitate meeting rigorous accreditation standards.

XSL•FO
RenderX

Future work may include combining orthopedic surgeon demand with the demand for specific orthopedic procedures. This would further develop our understanding of unmet orthopedic care needs in many states. For example, Cohen et al [12] measured the national public interest in the use of platelet-rich plasma for hip and knee osteoarthritis. In an analogous study with total knee arthroplasty and total hip arthroplasty, Cohen et al [13] also suggested that this data would be useful in making resource allocation decisions in states. Trends in the public interest may inform patient counseling, shared decision-making, and directions for future clinical research.

Limitations

Google was the only search engine used in this study, and it is possible that data from other search engines may yield different results, potentially affecting the generalizability of our findings. Additionally, while this study provides insights into regional trends, the absence of complementary data, such as hiring rates or job postings, limits the ability to draw definitive conclusions about the supply-demand dynamics in orthopedic surgery. The use of general search terms, such as “orthopedic surgery,” introduces potential biases, including the influence of surges of unrelated searches that may not reflect genuine patient interest. For example, during emergencies, the volume of orthopedic surgeries and searches for orthopedic services may not be as high. Our dataset does not control for this. Furthermore, online search data may not accurately represent underserved populations or individuals with limited internet access, introducing inherent demographic and geographic biases. These limitations underscore the need for integrating online search data with traditional epidemiological and workforce datasets in future studies to achieve more nuanced and representative analyses.

Conclusion

The supply of orthopedic surgeons varies greatly across the United States, with a higher demand for them in southern states. Demand and supply, represented by the RDI and orthopedic surgeon concentration, respectively, were strongly correlated negatively. States that were in the top quartile of residency programs (≥4 residency programs) exhibited a high demand for orthopedic surgeons. This study highlights the utility of the RDI as a scalable, data-driven tool for informing precision public health strategies. Beyond orthopedic care, the RDI framework could be adapted to other medical specialties, such as cardiology or general surgery, to identify underserved regions and guide resource allocation. For instance, the RDI can inform initiatives such as expanding residency programs, offering loan repayment or grant incentives, and deploying specialists to areas with high demand but limited supply. Additionally, if the RDI is correlated with specific procedures or conditions, such as knee surgeries in orthopedics or cardiac catheterizations in cardiology, it could be leveraged to target preventative public health initiatives. Examples include obesity reduction programs to decrease the need for joint replacements or smoking cessation campaigns to reduce cardiovascular disease prevalence. By integrating the RDI into broader public health frameworks, stakeholders can optimize care delivery, address disparities, and allocate resources more effectively across various health care settings.

Authors' Contributions

KM created Figures 2 and 3 and provided data for Table 2, FY and OI wrote the *Introduction* and *Methods*, and AQ wrote the *Results* and *Discussion* and created. KKN revised all text and provided mentorship. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

References

1. Moldovan F, Moldovan L, Bataga T. A comprehensive research on the prevalence and evolution trend of orthopedic surgeries in Romania. Healthcare (Basel) 2023 Jun 27;11(13):1866. [doi: 10.3390/healthcare11131866] [Medline: 37444700]
2. Fact sheet: orthopaedic surgeons will need to double total joint arthroplasty caseload to meet demand by 2050. American Academy of Orthopaedic Surgeons. 2023 Mar 7. URL: https://aaos-annualmeeting-presskit.org/2023/research-news/fact-sheet-orthopaedic-surgeons-will-need-to-double-total-joint-arthroplasty-caseload-to-meet-demand-by-2050/ [accessed 2023-06-01]
3. State population totals and components of change: 2020-2022. United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html [accessed 2023-11-29]
4. ERAS 2024 participating specialties & programs. Association of American Medical Colleges. URL: https://systems.aamc.org/eras/erasstats/par/display.cfm?NAV_ROW=PAR&SPEC_CD=260 [accessed 2023-11-29]
5. 2021 state profiles. Association of American Medical Colleges. URL: https://www.aamc.org/data-reports/workforce/data/2021-state-profiles [accessed 2023-11-29]
6. Orthopedic surgeons - explore. Google Trends. URL: https://trends.google.com/trends/explore?q=orthopedic%20surgeons&date=now%201-d&geo=US&hl=en [accessed 2023-11-29]
7. Pheister M, Cowley D, Sanders W, et al. Growing the psychiatry workforce through expansion or creation of residencies and fellowships: the results of a survey by the AADPRT workforce task force. Acad Psychiatry 2022 Aug;46(4):421-427. [doi: 10.1007/s40596-021-01509-9] [Medline: 34292538]

XSL•FO
RenderX

8. Cyr ME, Etchin AG, Guthrie BJ, Benneyan JC. Access to specialty healthcare in urban versus rural US populations: a systematic literature review. BMC Health Serv Res 2019 Dec 18;19(1):974. [doi: 10.1186/s12913-019-4815-5] [Medline: 31852493]
9. Ikpeze TC, Mesfin A. Interest in orthopedic surgery residency: a Google Trends analysis. J Surg Orthop Adv 2018;27(2):98-101. [Medline: 30084815]
10. Blau JA, Levites HA, Phillips BT, Hollenbeck ST. Patient demand for plastic surgeons for every US state based on Google searches. JPRAS Open 2020 Sep;25:88-92. [doi: 10.1016/j.jpra.2020.06.001] [Medline: 32904136]
11. Akosman S, Tran E, Rosenberg S, Pakhchanian H, Raiker R, Belyea DA. Patient demand for ophthalmologists in the United States: a Google Trends analysis. Ophthalmic Epidemiol 2024 Aug;31(4):349-355. [doi: 10.1080/09286586.2023.2273508] [Medline: 37885262]
12. Cohen SA, Zhuang T, Xiao M, Michaud JB, Amanatullah DF, Kamal RN. Google Trends analysis shows increasing public interest in platelet-rich plasma injections for hip and knee osteoarthritis. J Arthroplasty 2021 Oct;36(10):3616-3622. [doi: 10.1016/j.arth.2021.05.040] [Medline: 34172346]
13. Cohen SA, Cohen LE, Tijerina JD, et al. Google Trends as a tool for evaluating public interest in total knee arthroplasty and total hip arthroplasty. J Clin Transl Res 2021 Aug 26;7(4):456-466. [Medline: 34667892]

Abbreviations

AAMC: Association of American Medical Colleges
AAOS: American Academy of Orthopaedic Surgeons
ACGME: Accreditation Council for Graduate Medical Education
ERAS: Electronic Residency Application Service
GME: graduate medical education
RDI: relative demand index
RSV: relative search volume

Edited by E Mensah; submitted 03.09.24; peer-reviewed by F Moldovan, Q Tan; revised version received 08.12.24; accepted 12.12.24; published 31.01.25.

Please cite as:
Qiu A, Meadows K, Ye F, Iyawe O, Kenneth-Nwosa K
Quantifying Patient Demand for Orthopedics Care by Region Through Google Trends Analysis: Descriptive Epidemiology Study
Online J Public Health Inform 2025;17:e63560
URL: https://ojphi.jmir.org/2025/1/e63560
doi:10.2196/63560

© Abram Qiu, Kristopher Meadows, Fei Ye, Osasu Iyawe, Kenneth Kenneth-Nwosa. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 31.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of the Burden of COVID-19 Regulatory Reporting in a Small Independent Hospital and a Large Network Hospital: Comparative Mixed Methods Study

Yalini Senathirajah[1], PhD; David R Kaufman[2], PhD; Kenrick Cato[3], RN, PhD; Pia Daniel[2], MD; Patricia Roblin[2], MS; Andre Kushniruk[4], PhD; Elizabeth M Borycki[4], PhD; Emanuel Feld[1], MD; Poli Debi[2], MPH

[1]School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[2]Emergency Preparedness Division, SUNY Downstate Health Sciences University, Brooklyn, NY, United States

[3]Children's Hospital of Philadelphia, Philadelphia, PA, United States

[4]Health Information Science, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Yalini Senathirajah, PhD
School of Medicine
University of Pittsburgh
2607 Baum Blvd
Pittsburgh, PA, 15206
United States
Phone: 1 3476194021
Email: yalini@pitt.edu

Abstract

Background: During the COVID-19 pandemic in 2020, hospitals encountered numerous challenges that compounded their difficulties. Some of these challenges directly impacted patient care, such as the need to expand capacities, adjust services, and use new knowledge to save lives in an ever-evolving situation. In addition, hospitals faced regulatory challenges.

Objective: This paper presents the findings of a qualitative study that aimed to compare the effects of reporting requirements on a small independent hospital and a large network hospital during the COVID-19 pandemic.

Methods: We used both quantitative and qualitative analyses and conducted 51 interviews, which were thematically analyzed. We quantified the changes in regulatory reporting requirements during the first 14 months of the pandemic.

Results: Reporting requirements placed a substantial time burden on key clinical personnel at the small independent hospital, consequently reducing the time available for patient care. Conversely, the large network hospital had dedicated nonclinical staff responsible for reporting duties, and their robust health information system facilitated this work.

Conclusions: The discrepancy in health IT capabilities suggests that there may be significant institutional inequities affecting smaller hospitals' ability to respond to a pandemic and adequately support public health efforts. Electronic certification guidelines are essential to addressing the substantial equity issues. We discuss in detail the health care policy implications of these findings.

(Online J Public Health Inform 2025;17:e63681) doi:10.2196/63681

KEYWORDS

regulatory reporting; human factors; reporting burden; emergency response; COVID-19; hospital resilience; pandemic response

Introduction

Background

This research compared the impact of reporting requirements on a small independent hospital (SIH) that served a medically underserved population during the COVID-19 pandemic with a large network hospital that served the same city. The SIH, which had 198 staffed beds, was designated as a COVID-19–only hospital by the State of New York from March 28, 2020, to June 5, 2020. A large network hospital (LNH) with 11 campuses and greater resources was used as a point of comparison. Both institutions are in New York City (NYC), and their experiences during the same period were compared. This research has important implications for government and regulatory bodies, highlighting the need to adjust policies to

XSL•FO
RenderX

meet public health information needs without further straining an already overwhelmed health care system.

A core part of the mission of every public health agency is to consistently and methodically gather, compile, interpret, and disseminate data concerning the health status of the community [1]. Hospitals play a crucial role in supporting this mission by gathering and reporting data on their patient populations and institutional resources, which becomes particularly essential in public health crises. Approximately 96% of US hospitals have adopted certified electronic health record (EHR) systems, many with automated reporting for functions like public health surveillance [2]. This shift marks a major advance in the digitization of hospital records and the efficiency of data reporting, though challenges in interoperability still impede smooth data exchange across different institutions.

According to a 2019 Office of the National Coordinator report, around 50% of nonfederal acute care hospitals face electronic data exchange limitations with public health agencies [3]. Of these, 40% struggle with system interface issues that hinder data transmission, while 20% report difficulties because of the lack of a standardized vocabulary between hospitals and public health agencies, further complicating data sharing. Particularly, smaller, rural, and independent hospitals encounter challenges in extracting relevant data from EHRs for reporting purposes. The Office of the National Coordinator report found that 71% of hospitals experience significant challenges in public health reporting, which could limit public health agencies' ability to monitor and respond to disease outbreaks effectively.

A 2023 report by the Government Accountability Office similarly notes that while electronic health information exchange is widespread, smaller and rural hospitals continue to face obstacles [4]. The report summarized the results of a survey; small hospitals frequently rely on mail or fax, with 47.4% often sending and 54.5% often receiving information through these methods. In contrast, medium or large hospitals use mail or fax less frequently, with 30.3% often sending and 38.5% often receiving data this way. For regional, state, or local health information exchange organizations, 36.7% of small hospitals often send data and 23.2% often receive data, whereas medium or large hospitals are more engaged with these networks, with 56.4% often sending and 40.3% often receiving data. Similarly, for EHR vendor–based networks, 30.2% of small hospitals report sending data and 28.4% report receiving data, while medium or large hospitals show higher use, with 48.4% often sending data and 45.5% often receiving data through these networks. National health information exchange networks also show a gap, as 28.5% of small hospitals often send data and 23.3% often receive data, compared to 43.1% of medium or large hospitals often sending data and 36.5% often receiving data. Limited financial and technical resources, such as insufficient IT staffing, often hinder these hospitals' ability to sustain consistent electronic exchange [4].

Initiatives like the Trusted Exchange Framework and Common Agreement (TEFCA) aim to create a more unified data exchange approach [5]. TEFCA was established under the 21st Century Cures Act, which was signed into law in December 2016. Specifically, TEFCA aims to improve patient care, reduce health care costs, and enhance health care quality by enabling seamless data exchange among diverse health care entities. Although it is a promising initiative, participation remains voluntary and does not fully resolve these infrastructural challenges.

These findings were supported by a report from the Greater New York Hospital Association (GNYHA) [6], which highlighted the burdens experienced by hospitals during the initial months of the COVID-19 pandemic. The GNYHA report echoed the Office of the National Coordinator's findings, emphasizing the importance of reviewing data collection frequency and making necessary adjustments to strike a balance between public health requirements and hospitals' reporting capabilities. Since the onset of the COVID-19 pandemic, the response capacity of the health care workforce has been severely impacted, leading to acute stress reactions and illness among health care professionals [7]. These reactions were especially acute in small hospitals. Health information systems play a vital role in hospitals' efficient data collection and sharing of data. Such efficiency is valuable at all times, but particularly during a pandemic [8], because accurate and timely data on patient morbidity, mortality, and laboratory findings and institutional bed availability, personal protective equipment supply burn rates, and other concerns (eg, unit staffing and resourcing) are needed to guide decisions for a range of stakeholders involved in the public health response at city, state, and federal levels [8].

Objectives

The objectives of this paper are (1) to quantify the changing reporting requirements overtime during the pandemic, (2) to characterize the effect of reporting demands on a small independent safety net hospital with an LNH, and (3) to compare the effects of reporting requirements on the SIH and LNH.

Methods

Quantifying Reporting Burden

To illustrate the complexity of reporting, we quantified changes to federal reporting requirements over a 14-month period during the COVID-19 pandemic by examining the log of changes and requirements in the Health and Human Services Teletracking, a web-based system for reporting on hospital capacity. We also used interviews and documents to clarify requirements.

Participants

We interviewed various stakeholders in each institution, including hospital leadership (eg, chief executive officer, chief medical information officer, and chief nursing officer), directors of various hospital departments (eg, infection control, infectious disease, and respiratory therapy), emergency response management team, IT staff, and members of other services.

Settings and Data Collection

Stakeholders, such as the SIH, LNH, GNYHA, and NYC Department of Health and Mental Hygiene were interviewed using a semistructured interview approach using Zoom (Zoom Video Communications, Inc). Interviews were typically 60 minutes in length. Questions focused on decisions, work activities, tasks associated with each activity, resources used to

XSL•FO
RenderX

accomplish those tasks, and institutional goals and challenges to realizing those goals during the pandemic. See Multimedia Appendix 1 for the interview script.

Interviews were transcribed and coded using MAXQDA qualitative analysis software [9] by 5 coders working on different transcripts. Coding was reviewed by an independent coder for consistency and iteratively revised.

Data Analysis

We used the Systems Engineering Initiative for Patient Safety (SEIPS) model as a theoretical framework and as a basis to analyze the data [10-12]. Grounded in the macroergonomic work system model by Carayon and Smith [13], SEIPS offers a comprehensive approach to understanding the elements of a work system, making it highly relevant to health care environments. The model illustrates how work systems influence care processes, such as care pathways, patient journeys, and workflows, while linking patient outcomes with organizational and employee outcomes. This highlights not only patient safety but also worker safety and quality of working life. The SEIPS model consists of 5 key elements—person, organization, technology and tools, tasks, and environment—enabling its broad applicability [10]. It provides a robust framework for analyzing health care work, particularly under conditions of stress or disruption.

Central to our analysis are the concepts of resilience and adaptive capacity, both integral to the SEIPS framework. Resilience refers to a health care system's ability to adjust, recover, and continue functioning effectively in the face of disturbances, variability, or adverse events. Adaptive capacity refers to the ability of a system or its individuals to adjust to new circumstances and improve processes. The SEIPS model connects adaptive capacity to how the various work system elements interact to maintain safety and performance in dynamic conditions.

Once interviews were conducted, they were transcribed, cleaned, and coded using a schema based on the SEIPS 2.0 models [10]. The SEIPS model and coding scheme was used to identify the critical aspects of the organization and its associated decision-making, tasks, and experience, which facilitated or challenged the organization's ability to maintain resilience. The codings schema included 13 top-level codes based on a modified version of the SEIPS model developed by Carayon with additional concept codes specific to this context [11]. The focus of this paper is predominantly on infection control and reporting burden.

Ethical Considerations

All procedures were approved by the Downstate Medical University Institutional Review Board (1638083-7). Anonymized data was stored on encrypted servers or hard drives in locked offices, and referred to only by a code. No compensation was given to interviewees.

Results

Overview

We interviewed 51 people, including 32 at the SIH, 16 at the LNH, and 3 with 2 external agencies, GNYHA and the NYC Department of Health and Mental Hygiene.

Quantifying Reporting Burden

Reporting clinical and public health data such as numbers of tests, infections, hospitalized patients, and deaths from COVID-19 were mandated by NYC at city, state, and federal levels, via several different agencies and jurisdictions. Table 1 shows the data and agency and provides examples of reporting requirements. The timeline in Figure 1 shows how these have changed over time for the federal Human and Health Services Teletracking system. The graph quantifies burden on reporting staff, showing total changes that reporting staff must handle for each time period. This includes new requirements being added, the total number of data points, and data points which were changed (which also necessitates work even if requirements were removed), as well as total reportable elements. All these together constitute burden. The data were gleaned from the interviews, publicly available documents, and guidance from the emergency response management team that oversaw the process.

The bulk of reporting was to the NYC and New York State departments of health. The schedules varied from several times a day to intermittently or as requested. During the first few months of the COVID-19 pandemic, the reporting requirements increased dramatically. At the SIH, 17 departments were engaged in the reporting process. Nursing and pathology were each involved in 7 reports. The New York State Department of Health reporting was more demanding than the others, and as we discussed, infection control was the most heavily impacted. There were sharp increases and changes across agencies. At the beginning, there were 5 reports which later increased to 29.

Figure 1 includes the total number of data elements, changes in reporting requirements, and other changes. Each of the columns represent a date between April 27, 2020, and June 10, 2021. The columns reflect the 23 dates when a change was made to reporting in that time interval. Of the total 23 changes, 18 changes occurred in 2020, and 6 changes were in July (columns 5-10). The stacked plot shows the total number of data elements to which reporters had to pay attention. On a couple of occasions, data elements were removed, and this was also reflected in the changes. Any change, including removal, could have serious effects on data gathering and processing workflows and personnel. The federal data elements increased sharply at the beginning from 110 to around 170, which coincided with the introduction of the Teletracking system and fluctuated after that.

XSL•FO
RenderX

Table 1. Reporting requirements for New York City institutions.

Agency	Data	Submission schedule	SIH[a] responsible departments
NYCDOH[b], NYS[c]	Patients with COVID-19 and their bed summary	Daily by 4 PM (later 1 PM)	Nursing, infection control, and risk management
NYCDOH, NYS	Data for COVID-19 pandemic bed increase directive	Onetime request	Emergency management and admitting
NYCDOH, NYS	Fatalities (particularly related to COVID-19)	Within 24 h of death	Pathology, infection control, and nursing
NYCDOH, NYS	PPE[d], split ventilators, and staffing	One time request	Central supplies, respiratory therapy, and nursing
NYCDOH, NYS	COVID-19 death certification	Within 24 h of death	Pathology and admitting
NYCDOH, NYS	Dialysis needs: staff, supplies, and equipment	Onetime request	Dialysis, central supplies, and nursing
NYCDOH	Pediatric patients with COVID-19	Daily	Nursing, pediatric infectious diseases, admitting, and risk management
NYCDOH	NYCDOH syndromic surveillance ED[e] data	Daily	IT and emergency department
NYCDOH	Pediatric multisystem inflammatory syndrome	One time request	Nursing, pediatric infectious diseases, admitting, and risk management
NYCDOH	Syndromic surveillance data and patient ID	Daily	IT and emergency department
NYCDOH	Temporary ED data	One time request	Emergency department
NYCDOH	NYC[f] communicable diseases preparedness	One time request	Emergency management and infection control
HHS[g]	COVID-19 test data from in-house and commercial laboratories	Daily by 5 PM	Pathology and microbiology laboratory
OASAS[h]	COVID-19	One time request	Pathology and microbiology laboratory
GNYHA[i]	Sit Stat daily report	Daily by 11 AM	Emergency management
GNYHA	Inpatient psychiatric census	One time request	Psychiatry, infection control, and nursing
GNYHA	Nursing home EO[j] impact	2 times/wk	Emergency management
OCME[k]	Hospital morgue census or availability	Daily by 3 PM	Pathology and laboratory
AHA[l]	Hospital laboratory data	Daily	Pathology and microbiology laboratory
HANYS[m]	COVID-19 extraordinary hospital and health system cost	One time request	Emergency management, hospital administration, hospital finance, and clinical services
National Guard	Bed capacity and census	Daily every 12 h	—[n]
CMS[o]	Average cost for specified outpatient drugs	By May 1	Pharmacy, hospital administration, and hospital finance
FEMA[p]	Number of COVID-19 tests and type, antibody test, antibody, and PCR[q]	Daily report due by 5 PM	Pathology and microbiology laboratory

[a]SIH: small independent hospital.
[b]NYCDOH: New York City Department of Health and Mental Hygiene.
[c]NYS: New York State Department of Health.
[d]PPE: personal protective equipment.
[e]ED: emergency department.
[f]NYC: New York City.
[g]HHS: Health and Human Services.
[h]OASAS: Office of Addiction Services and Support.
[i]GNYHA: Greater New York Hospital Association.
[j]EO: emergency office.
[k]OCME: Office of the Chief Medical Examiner.
[l]AHA: American Hospital Association.
[m]HANYS: Health Administration of New York State.
[n]Missing data.

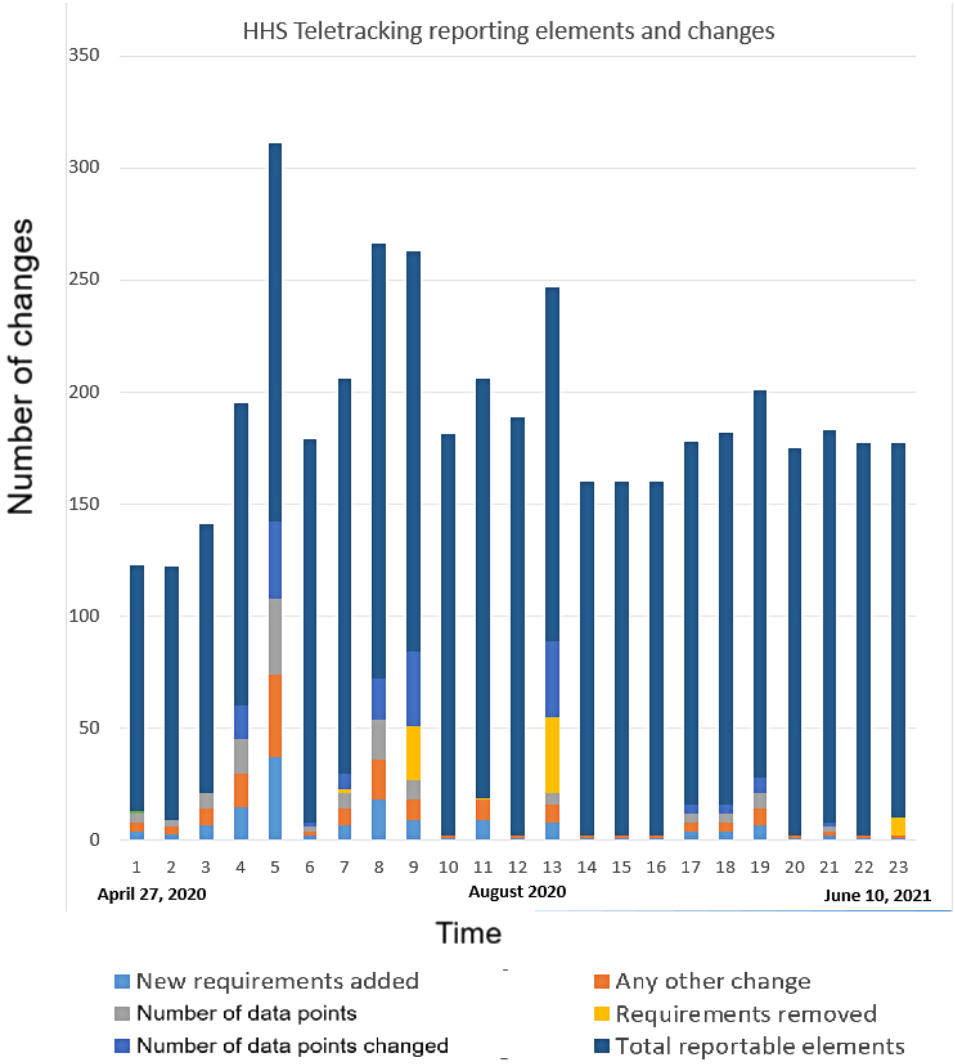
XSL•FO
RenderX

[o]CMS: Center for Medicare and Medicaid Services.

[p]FEMA: Federal Emergency Management Administration.

[q]PCR: polymerase chain reaction.

Figure 1. Changes in federal reporting requirements via the Department of Health and Human Services (HHS) Teletracking service.



Definitions changed periodically and created significant stress for those involved in the reporting process. The number of staffed beds versus “available beds” was an example of how definitions were a source of miscommunication. In addition, different agencies used different variables for the same measure regarding bed availability, surges, surge capacity, and so forth. Sometimes, the variables were redefined midstream, necessitating a change in data gathering. It was also problematic that the SIH had 376 total licensed beds but only 198 were staffed. This resulted in discrepancies as it is used as a denominator to estimate available capacity (eg, beds available immediately for an emergency), causing consternation and disputes with agencies.

Interviews

The interviewees included hospital leadership, medical directors, frontline clinicians, directors of laboratory, nursing, infection control, and other services; their technical staff; those in charge of facilities and logistics; emergency staff; and hospital staff.

Table 2 includes concepts mentioned in the same context (eg, coding segment) as an information report. The state government was the most focally discussed concept in this context (ie, 111 instances). It is the primary recipient of mandatory reporting and posed issues for individuals and institutions responsible for the reporting. Federal agencies were mentioned in 62 instances. Information gathering was seen as a rate-limiting factor in the reporting process because of interoperability and data extraction problems. The interviewees often discussed the arduous nature of the reporting process and how it constituted a burden. Infection control reporting requirements emerged as a central focus. The data gathering involved a collective and often manual process. However, the infection control and infectious disease departments shouldered a disproportionate burden. Particularly, the infection control department had substantial difficulties executing their routine hospital surveillance tasks during the pandemic.

XSL•FO
RenderX

Table 2. Concepts co-occurring in the same segment of text (typically a sentence or two) with the code "information report" (n=423).

Variable	Instances, n (%)
Agencies: state	111 (26.2)
Agencies: federal	62 (14.6)
Automated vs manual	46 (10.9)
Information gathering	40 (9.5)
Burden	33 (7.8)
Infection control	29 (6.9)
COVID-19 tests	24 (5.7)
Communication	15 (3.5)
Deaths	15 (3.5)
Patient census	15 (3.5)
Technology tools	13 (3.1)
Agencies: city	10 (2.4)
Work processes	10 (2.4)

Emergent Themes

Differences Between the Two Hospital Systems

There was a stark difference in the impact of reporting requirements on staff at each institution. The LNH had a 50-person analytic team that handled reporting tasks, including data collection and submission. The involvement of clinical staff in reporting was limited to approval and coordination. Reporting was not as much a focal point in LNH interviewees' responses, even though the interviewers asked about it specifically. By contrast, reporting was a cross-cutting theme in the SIH interviews, with 90% (29/32) of interviewees mentioning it, often in the context of barriers and burdens.

Time Constraints, Penalties, and Other Factors Affecting the Task

Typically, new reporting requirements for NY State were first conveyed to institutional contacts at 8 AM with a required deadline of 1 PM the same day. Penalties for missing reporting deadlines posed challenges for the SIH staff, many of whom also faced increased clinical demand during the COVID-19 pandemic which was further complicated by staff shortages. Not only could there be a lack of clarity around new data elements' definitions, but also, at the SIH, an added reporting requirement could mean nontrivial changes to workflows: conversations with staff, data collection via phone calls or observation, manual extraction from systems without automated exports, tabulation in spreadsheets, and manual entry into reporting forms. Furthermore, the continuous changes made it difficult to develop stable data gathering and workflow processes.

The excerpts in Tables 3 and 4 illustrate the challenges experienced by various personnel involved in regulatory reporting.

As seen in Table 3, the comments from SIH personnel (ie, quote numbers 1-6) highlight the challenges and urgency of meeting reporting requirements and the problems associated with locating and extracting the needed information (ie, quote number 3). Moreover, the burden of reporting may have diminished the resources needed for critical care decisions (ie, quote number 5). The ability to display data in dashboards and use it for predictive modeling was also a major need (ie, quote number 6).

In contrast, the LNH had ample resources from the baseline before the pandemic. However, they also faced challenges associated with the distribution of responsibilities and resources (ie, quote number 7) and experienced the pressures of needing to report in a timely fashion (ie, quote number 8). The dedicated analytics team managed data aggregation and reporting, largely eliminating the burden on other staff. They also had a centralized data warehouse that automatically updated, reducing the need for manual searching and compiling data for reports. Their electronic medical record generated reports that could be accessed by anyone, not just the IT team.

Table 5 shows the distilled data from interviews and documents and compares the reporting process at the SIH and LNH.

There are vast differences in the technical capabilities between the SIH and LNH, typical of the gap between wealthy private hospital networks and small stand-alone hospitals, as seen in Table 5. The manual information gathering processes at the SIH and the fact that the LNH has 50 people involved in data aggregation and analytics, starting from an already highly automated aggregation system, led to profound differences in the reporting workflow.

Table 5 calls attention to several issues pertaining to the processes involved and what might be needed to improve overall. Automation levels, which directly affect burden, are only part of the problem.

XSL•FO
RenderX

Table 3. Illustrative quotes.

Quote number	Quotes from the SIH[a]	Codes
1	"For the Department of Health in New York State, they want the report at a set time. If you don't put it in at that time, you get penalized, and it's a huge fine for the hospital. So now we have to meet all the mandates together with meeting your primary responsibilities before COVID, right?" [Infection control leadership, SIH]	• Situated resilience: respond or respond to task demands • Task: info report • Task: burden • External environment: state
2	"We have to extract it, so it's not a 100% foolproof process. Sometimes you have to do manual extraction, which is very tedious, and that takes the whole time. So between nine to one-o-clock, that's what we're working on." [Infection control leadership, SIH]	• Task: information gathering, person or domain: nursing • Task: burden • External environment: state
3	"We have a number of machines that are not interfaced yet with our EMR[b] for one technical reason or another, and so instead, the technician is responsible for manually entering each result..." [Laboratory administration, SIH]	• Tech tools: automated • Internal environment: technology
4	"Nobody's analyzing the cases as they come into the hospital. What's their O2 stat in an aggregated way? How sick are they? Are they having fevers? Are they getting discharged? How long has it taken for them to be discharged? I would be really interested in that sort of thing to understand the complexion of the epidemic." [Infectious disease leadership, SIH]	• Structural resilience: capacity to monitor patients • Tasks: information gathering • Info type: patient census • Potential solution
5	"We didn't even know what it meant to be anointed, a COVID-only hospital. Would we get busloads of people? We immediately had to draft scenarios. We have our main hospital and a surgery center a couple of miles away, and we drafted two different models for each site. What if we were a COVID-only hospital at 50% critical care level and then 80% critical care level; what would we need?" [Clinical services, SIH]	• Person domain: critical care or ICU[c] • Tasks: predict • Adaptation: planned or anticipated • Systemic resilience: anticipate demands or capacity misalignments
6	"It'd be also nice to have longitudinal data, of course as well, so taking a look at how things have evolved over the past week over the past month, over the past two months, just to see how those things have changed over time, so I've heard mention of a dashboard. But I haven't seen it implemented yet." [Laboratory administration, SIH]	• Person domain: PPEd supplies • Tech tools: potential solution • Tech tools: dashboard visualization • Structural resilience: monitor or monitor opportunities
7	"[D]uring the beginning I was intubating patients at a rate of one per hour, saving lives. Then I was taken off to save lives by filling out spreadsheets." [Emergency physician, SIH]	• Person domain: emergency medicine • Respond: respond to task demands • Adaptation: reaction or unanticipated

[a]SIH: small independent hospital.

[b]EMR: electronic medical record.

[c]ICU: intensive care unit.

[d]PPE: personal protective equipment.

Table 4. Illustrative quotes.

Quote number	Quotes from the LNH[a]	Codes
8	"[Named person] was one of the people responsible for filling that out early on, and it was one, trying to figure out the definitions, was it accurate, did her... she had to get it quickly approved through the chief operating officer before it would go out centrally, and at the same time, she's trying to build a two hundred-bed field hospital in a week." [Nursing leadership, LNH]	• Tasks: info report • Tasks: burden • Person domain: executive • Person domain: nursing
9	"So there were challenges during that time period that... this was sort of like, 'Why is this necessary? We're just trying to survive here.' And yet, it was necessary to understand what was happening in the State of New York, so when you think about it logically, of course, we needed to provide data about what was happening in our hospitals, but at the time when you're choosing between... are you gonna create a bed to try to take care of another patient, or are you gonna slow down enough to present accurate data? It was annoying because it couldn't work." [Pharmacy leadership, LNH]	• External environment: State government • Monitor: monitor task demand or capacity misalignment • Tasks: info report • Tasks: burden

[a]LNH: large network hospital.

XSL•FO
RenderX

Table 5. Comparison of the reporting processes at the SIH[a] and LNH[b].

	SIH	LNH
HHS[c] daily report (remdesivir data, monoclonal antibody, COVID-19 positives, pediatrics, and bed capacity)	• Rely on department collection of info into Microsoft Excel • Small differences in definitions and strict deadlines • Remdesivir data from pharmacy • Reliance on paper and Microsoft Excel	• NSHN[d] first report March 30, 2020—3 days to prepare, ICU[e], ventilator counts, 15 nurse team, coordinated by senior vice president or chief operating officer
HERDS[f], NYS[g]	• Daily manual inventory of intake or outflow • Need for real-time inventory amplified by crisis • PPE[h]: central supply email and Microsoft Excel; phone calls used to reconcile information (eg, sizes and models) • Hard to track supplies from orders, expected time of arrival, delivery, and use • Historical data kept in spreadsheets • No centralized place to share equipment data • Coordinated on needs with procurement, central sterile, infection control, disaster management, and facilities management • Admissions of patients who tested positive for COVID-19 infection from EMR[i] in mornings; internal isolation list emailed to leadership	Initial: • Manual data entry for each patient; requested extra time as this was impossible during high loads • Infection preventionists and regional hospitals submitted manual reports via phone, email, and fax, populated an initial database and Microsoft Excel • Used for accuracy tracking to develop automated reports, suspect cases were hard to capture • Rapid multiple changes and penalties for not submitting in a few hours • State required someone available to answer questions 24/7. Manually intensive Later: • Automated via laboratory, centralized reporting from 2 laboratories, saved in Cerner, generated reports 4 times per day.
HERDS: Demographics of patients with COVID-19 (age, employment and essential worker status, how they get to work, and where they were admitted from): deaths. Initiated several months later.	• Need for HERDS report to be submitted by 1 PM (working on it from 9 AM to 1 PM). • Manual extraction and work with IT to automate or integrate into admission records so that these things are noted by nurses. • Lapses where nurses forget to ask these questions. HERDS has to be manually input. Someone else has to attest to it before submission. • Deaths by COVID-19: deaths and mortalities via an IT-generated report in Microsoft Excel emailed to attester. • COVID-19 discharges: from an IT-generated report emailed to attester. • Compare across days with the internal isolation list to ensure that they did not miss any discharges.	• HERDS required 2-h turnaround (later 3 h) daily. • Analytics dept consisting of clinicians do registry reporting. • Both centralized and decentralized—some campuses had capacity, others did not. • End of year: 90% HHS reports automated. • Medication reporting: by pharmacy, 13 hospitals via email to one person daily, (then weekly, monthly, over) Vitals, patients, inventory, no inventory system, spot-checking, manual decrementing, 2 h per day for pharmacy person. “Paralyzing.” • Report via a smartsheet application. Vague definitions initially. [staff] filed, had to get it quickly approved by the chief operating officer, while building 200 bed field hospital in a week. • HERDS reporting now done by 20-person team.

[a]SIH: small independent hospital.

[b]LNH: large network hospital.

[c]HHS: Human and Health Services.

[d]NSHN: National Healthcare Safety Network.

[e]ICU: intensive care unit.

[f]HERDS: Health Electronic Response Data System.

[g]NYS: New York State Department of Health.

[h]PPE: personal protective equipment.

[i]EMR: electronic medical record.

Ability to Change

Although manual processes are time-consuming and manually coordinated, process changes cannot be easily implemented; unlike the automated data extraction process that takes 10 minutes, there is far less leeway for errors or further changes. Because manual processes require the availability of specific personnel, changes may require additional coordination and sometimes staffing and may not be feasible. Staff retraining may be needed.

Institutional Self-Knowledge

Our studies revealed that self-knowledge by the institution is also lacking—our research served to inform staff internally of who was doing what in other areas outside their own, enabling us to pull together an overview and details hitherto unknown by administrators (ie, “organizational situational

XSL•FO
RenderX

self-awareness"). Again, this can be a function of lack of IT resources and knowledge of how to use them; such overviews are commonly made in IT process changes where the "as is" and "to be" workflows are explicitly illustrated using convention-based diagramming and planning tools.

Effect on Clinical Work

The current processes took clinical staff away from urgent clinical duties, and the effects of this on actual patient outcomes are difficult to quantify, but concerning, particularly for this already underserved population.

It is notable that even common tools, such as shared drives, were not necessarily used (eg, "no central space to share equipment data"). Implementing changes to incorporate digital tools is further hampered by the need to involve IT staff who are overwhelmed by daily duties and emergency needs and unable to consider further process changes even though it would relieve burdens in the long run. Clinical staff made use of external digital tools, such as WhatsApp for coordination as these are freely available, can be set up by clinicians, and meet security requirements. This illustrates what is needed to be useful in future.

However, there were also similarities and common challenges. Neither institution adequately planned for a pandemic of this magnitude. Both had to change rules and procedures and make numerous adaptations. The SIH also had to deal with state and other mandates arising from its status as a COVID-19–only hospital and as part of the state hospital system. This meant that it could not refuse patients with COVID-19 but did not receive extra compensation. There was a significant revenue loss because of canceled surgeries and other procedures. This burden affected all hospitals, but the ones which were resource-limited were more greatly impacted.

Discussion

Major Findings and Implications

Differences Between the Two Hospital Systems

The COVID-19 pandemic placed a great burden on health care institutions throughout the country. SIHs disproportionately felt the burden. Regulatory reporting is an essential aspect of public health and takes on greater importance during a crisis like the COVID-19 pandemic. Unfortunately, the reporting requirements added to the burden and the great stress these institutions were already experiencing, with a more significant impact on SIHs.

Health IT inequities have been well-documented, and their detrimental impact on health disparities, exacerbated by the COVID-19 pandemic, has been noted [14].

We found that the lack of health IT infrastructure and staffing imposed significant demands on clinical staff and rendered patient care more difficult. We compared the impact on an SIH with a much larger and better-resourced network hospital. Although both endured substantial challenges during the COVID-19 pandemic, the smaller institution disproportionately felt the effect. The larger one had sufficient staffing and the capabilities to support evolving reporting needs, supported by greater automated data extraction capabilities. The findings are consistent with an ONC survey of hospitals conducted before the COVID-19 pandemic, which suggested that many hospitals were ill-prepared [3].

Moreover, the distribution of federal COVID-19 relief funds to hospitals often failed to prioritize the communities most in need [15]. However, it is the institutional inequities, particularly concerning disparate IT infrastructure and automation capabilities, that are deeply concerning and pervasive. While variations between the SIH and the LNH were expected, the differences were more profound and had far-reaching consequences. To establish a truly equitable public health infrastructure that meets the requirements of a modern health care system, adequate resources must be allocated to support the implementation of mandates. This includes funding for training programs and establishing measures to address the disadvantages faced by SIHs. IT has become the backbone of health care, and it is crucial for public health agencies to acknowledge that hospitals with limited IT resources will struggle to fulfill reporting requirements. When feasible, demand should be adjusted or, at the very least, temporal demands should be reduced to accommodate the hospital reporting capacity. Recognizing the disparities in IT capabilities and addressing them is a crucial first step toward building a more equitable health care system. By acknowledging and addressing the challenges faced by resource-limited hospitals, public health agencies can foster a more inclusive and effective reporting framework that supports the needs of all health care providers.

Time Constraints, Penalties, and Other Factors Affecting Tasks

Punitive Measures Compounding Hospital Stress and Duress

What made reporting requirements more stressful was the punitive measures of some agencies. A report filed a few minutes late could result in a letter sent to the hospital president with the threat of disciplinary action. Different agencies sometimes created contradictory mandates. For example, the state threatened a US $1 million fine for giving vaccines to those outside the proper queue category, while the city threatened fines if all the vaccines were not used by a certain date. Because of the complex and difficult logistical and reporting process, institutional managers at the SIH paused the vaccine distribution for some time. The punitive government rules compounded the stress and burden of manually aggregating data, deciphering sometimes ambiguous requests, confusing data formats, and delivering a report by 1 PM. It also diminished available resources for patient care. For example, some clinical staff in infectious disease spent around 2 hours a day gathering, aggregating, and verifying information. Other clinical personnel at the SIH were also taken away from clinical duties to accommodate reporting. Notably, there was a significant increase nationally in hospital-acquired infections in 2020 compared to 2019 [16].

Health care personnel must be more involved in making the rules and consulted on the impact of rule changes. The agencies' determination of "need-to-know" needs to be reconciled with the impact of sudden changes to reporting requirements. In addition, the need for strict compliance and enforcement through

XSL•FO
RenderX

punitive measures needs to be tempered by the impact of such measures on hospital staff.

Government Mandates Often Presume a Level of Existing Technology

The United States cannot expect resource-limited institutions to follow all unfunded mandates. This is most evident in circumstances, such as a pandemic, but it is the state of affairs at all times. Furthermore, it is reasonable to speculate that *health IT inequities contribute to health disparities in patients.* What effect does a lack of automated data collection and aggregation have on the data collected for patient management? How does this affect the equity of quality assurance, needs assessment, and research? This merits further research.

The effect on clinical work is one of the most deeply concerning results, with clinical staff being removed from caring for patients, and the systemic issues contributing to a decline in service as well as new needs, innovations, and workarounds.

Future Needs and Directions

The Need for Regional Data Sharing

A recurrent theme was the need for centralized websites where regional hospitals could list their available supplies and other needs. No means were provided to leverage the data reported to public health agencies to foster cooperation regionally to ensure supplies, staffing, medications, and predictive modeling of surges, which are naturally affected by regional infection rates. The SIH, for example, often experienced a 2- to 3-week lag in COVID-19 pandemic waves compared to neighboring hospitals, and regional data would support efforts to make anticipatory changes. On a related note, it is often difficult for hospitals to repurpose the data generated for reporting purposes to address other hospital needs, such as quality control. Regional data sharing could serve to coalesce data and render it more reusable. In addition, public health agencies need to provide more granular data on COVID-19 cases, deaths, and vaccinations available to allocate resources to mitigate disease spread [17].

Need to Gauge Hospital IT Readiness for the Pandemic

Currently, NYC funds both large and small hospitals for infrastructure, and there is competition for these resources, with large networks arguing that they have greater needs. However, it is clear from our study that basic resources are not available; the 2 types of institutions start from very different baselines. Individual nonnetwork SIHs lack the minimum resources in IT infrastructure and personnel to function efficiently; rather, they are reliant on the heroic efforts of staff. To establish whether a hospital or health system has instances of inequity in IT resources, it may be useful to formulate a rubric of questions and a scaling system analogous to the Health Information Management Systems Society 7-Level Health IT Maturity Model [18].

Need for Centralization of Data Aggregation Within the Institution

One difference between the SIH and the LNH was that the LNH data collection was centralized, reducing the burden of multiple people doing things manually and increasing the consistency and reliability of the data. Resource-limited institutions could benefit from resources to set up similar centralization, including 1 or 2 personnel dedicated to the reporting task.

The Need for Health Care Professionals to Insist on Independent Control and Sharing of Their Data With Commensurate Health IT Development

In the competitive landscape where public reputation, patient experience narratives, and scoring (eg, Press Ganey and Hospital Consumer Assessment of Healthcare Providers and Systems scores) have financial consequences, institutions are wary of sharing data. There is a concern that exposure might put them in a bad light, provide a means of competitive attack, decrease their possible benefits from exclusive data access for research or startup activity, or have other consequences. This was heightened during the COVID-19 pandemic. For example, normal research data sharing was often suspended. There is also a lack of reciprocation in which public health agencies make the submitted data with or without additional analyses available to the hospitals. For much of the pandemic, institutions reported to governments without timely aggregated access to these data not "knowing what they already know." It may help to set up a health care institution–owned network for data sharing and aggregation outside the government, with appropriate governance, to help institutions know information vital for their operation.

A corollary of this need is the fact that new technologies permit much greater creation and control of functionality by end users. The advent of no-code and low-code tools and "end-user design/programming" in other areas has demonstrated how these can sharply reduce costs and time taken while providing the needed functionality [19]. A further issue is that the control of technology by clinicians on the front lines is likely to be a better fit to the tasks and contexts involved and, if it has a suitable architecture, more easily able to meet the rapidly changing needs found in emergency situations. A greater focus on ensuring such suitable tools are incorporated into existing systems will help ensure that they are available when needed. The SIH was recently cited by Joint Commission on Accreditation of Healthcare Organizations as having one of the best-prepared emergency response frameworks, a critical part of which is the principle that emergency processes are in place and rehearsed as part of normal functions, not suddenly implemented during a new emergency.

New ways of capturing data or streamlining data capture are also needed. For example, instead of having nurses collect numbers of patients with COVID-19 in the ward on paper and then transferring it to Microsoft Excel, apps where such counts could be entered on a phone and automatically put into aggregation and visualization tools available to all on the web would be much more efficient. Voice input can be easier and faster.

Therefore, to "get from here to there" requires careful thought about the steps needed to feasibly integrate digital tools into the institution to aid emergency response.

This may require a 2-pronged simultaneous approach as follows: first, conducting tool selection and acquisition by those with

XSL•FO
RenderX

the autonomy to use them on their own (as the clinical services did with WhatsApp), and second, planning steps for which IT staff are involved, with additional resources. Generalizable tools, such as the ability to post on the internet with user-selected granularity of who can view, edit, acquire, or change information, may help meet needs while being sensitive to our findings such as institutional reluctance for sharing data with competitors.

Need for a Robust National Data Infrastructure

Despite recent large US investments in health IT, the COVID-19 pandemic revealed longstanding gaps and ongoing challenges in the local, national, and global public health information systems and data infrastructure and a concurrent and underlying data and information crisis and need for mitigation strategies [20,21]. The US public health system is overdue for a real-time, technology-driven surveillance and reporting infrastructure to respond adequately to public health emergencies [22]. Tracking daily hospitalization data at a national level would be a major step forward in quantifying the hospital system's real-time impact. Many public health agencies and hospitals continue to rely on manual processes to gather, collate, and submit data [23]. In addition to inefficiency, it often results in incomplete and missing information and a compromised ability to predict and map infection surges [23]. It is also well-documented that there was a slow uptake of findings from clinical trials. Dron et al [24] argue that this is the time to advance scientific cooperation and shift the clinical research enterprise toward a data-sharing culture to optimize our response in the service of public health.

The Importance of Equitable Universal Data Capabilities and the Need for EHR Certification Guidelines

The widespread use of EHRs in our health care system should provide fertile ground for more real-time standardized electronic EHR-based public health reporting. In their review of EHR-based surveillance systems, Aliabadi et al [25] identified 6 challenges to EHR-based surveillance as follows: policy and regulatory, technical factors, managerial factors, standardization, financial factors, and data quality. On the federal level, the EHR certification guidelines, alongside the interoperability framework by TEFCA, serve as both incentives and requirements that could enhance EHRs' capacity to support near real-time public health reporting. From a technical perspective, the 21st Century Cures Act addresses the critical technical issues of interoperability and data sharing [26]. However, as has occurred with similar initiatives, well-resourced institutions are likely to meet these requirements, while others, such as the institution we studied, may lack funding and struggle to comply. This could lead to further digital disparities with consequences for future emergencies. Establishing these requirements as part of permanent EHR certification guidelines, aligned with TEFCA's interoperability goals, would help address these issues in the long term.

Despite advances in EHR interoperability, many health care organizations still face significant challenges. These stem from the use of proprietary data formats and varied terminologies across different systems. The absence of a unified data model complicates efforts to normalize and exchange information seamlessly, leading to inefficiencies in reporting and analysis. In addition, the current certification process does not mandate for implementation of the most up-to-date public health–based interoperability standards (eg, Health Level 7 and electronic case reporting [eCR] Fast Healthcare Interoperability Resources [FHIR]).

The US Centers for Disease Control and Prevention has implemented a rule requiring eCR [27], but eCR is not currently mandated. As the regulation stands, institutions have until January 1, 2026, to implement eCR FHIR or demonstrate an alternative working method. There is a recommended FHIR app that can be used to implement this as well as technical and process advisories. However, as has occurred with similar initiatives, the likely outcome is that well-resourced institutions will be able to meet these requirements, while others such as the institution we studied, which do not benefit from funding for tribal health settings and Federally Qualified Health Centers, may struggle to comply. This is likely to lead to further digital disparities with consequences for future emergencies. Our study supports and highlights the great need for financial support for all low-resourced settings to make eCR truly national. Establishing this as part of permanent EHR certification guidelines would also help address the issues in the long term.

Potential Features and Benefits of Automated Reporting

Automated reporting plans would build on the current eCR rule, including automated extraction, consistent data models, the ability to rapidly specify new data elements and sources (not currently included), rules outside the current triggers that focus on numerical codes, medications and tests, and a complete end-to-end testing in situ, including the personnel and simulated emergent conditions. This would have benefits including (1) rapid response and surveillance of emergent conditions; (2) little need for programmer involvement during crises; (3) complete data from all sites, giving a truly accurate picture of what is happening as it happens; and (4) far more equitable health system development, with the release of personnel currently highly burdened by reporting requirements. Annual influenza season reporting could be used as a practice initiation exercise to refine requirements and test systems.

EHR certification requirements or guidelines are essential to long-term equity and consistency. As long as functionality is optional and unfunded, lower-resourced settings are unlikely to afford compliance, as we see with the current FHIR mandates, which were much delayed in enforcement. Making eCR a certification requirement also puts the onus on the vendor, who has the capacity and funds to implement it. Pandemic emergency response is critical for the nation, and it would be unwise to leave the current "holes" in our defenses, given the potential exponentially rapid spread of some conditions. A guarantee that functionality is in place before emergencies arise would help ensure response feasibility, rather than having the mad scramble that usually takes place.

XSL•FO
RenderX

Limitations of the Study

The interviews for this study were conducted approximately 6 to 10 months after the initial wave of the COVID-19 pandemic. It is important to acknowledge that memories during this time frame may be prone to inaccuracies or gaps. However, it is fortunate that despite this potential limitation, the interviews yielded substantial convergence of information, which enhances the credibility of our conclusions. It is worth noting that some potential interviewees declined participation because of their reluctance to revisit the traumatic nature of their experiences. This aspect should be taken into account when considering the representativeness of our findings.

While the generalizability of our findings may be somewhat limited because of the data collection being focused on 2 specific settings, it is noteworthy that when we shared our findings with affiliated hospitals, we received feedback indicating that many other institutions encountered similar challenges. This suggests a broader applicability of our findings beyond the settings studied. It is crucial to continue gathering evidence from various institutions and contexts to gain a comprehensive understanding of the widespread problems faced by health care providers during the COVID-19 pandemic.

Conclusions

Pandemic-related reporting imposes considerable burdens on all institutions. SIHs experienced a greater burden given that they may have had to rely on manual data collection processes. This may pull clinical staff away from critical clinical duties, affecting care. Greater attention, funding, and development of ITs and training and the setup of dedicated personnel for reporting and analytics may decrease this burden. Using a formal assessment rubric to determine the state of health IT readiness across institutions would assist in comparing institutions for the determination of needs. Greater regional coordination and addressing intangible issues in emergency response are also needed to present a better equipped and more resilient institutional future.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to privacy and institutional review board requirements but are available from the corresponding author on reasonable request.

Conflicts of Interest

Authors YS and DRK are Associate Editors for *JMIR Human Factors*; AK is the Editor-in-Chief for *JMIR Human Factors*; EB is the Editor-in-Chief for *JMIR Nursing*, and Associate Editor for *JMIR Human Factors*; KC has been a guest editor for *JMIR Nursing*. These authors had no involvement in the peer review, editorial review, or any aspects of editorial processing of this manuscript.

Multimedia Appendix 1
Health IT COVID project: interview questions—activity interview.
[DOCX File , 20 KB - ojphi_v17i1e63681_app1.docx]

References

1. Institute of Medicine (US) Committee for the Study of the Future of Public Health. The Future of Public Health. Washington, DC: National Academies Press; 1988.
2. Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of electronic health record systems among U.S. non-federal acute care hospitals: 2008-2015. ONC Data Brief 2016 May;35:1-11 [FREE Full text]
3. Richwine C, Marshall C, Johnson C, Patel V. Challenges to public health reporting experienced by non-federal acute care hospitals. The Office of the National Coordinate for Health Information Technology. 2019. URL: https://www.ncbi.nlm.nih.gov/books/NBK609329/pdf/Bookshelf_NBK609329.pdf [accessed 2024-04-29]
4. Electronic health information exchange: use has increased, but is lower for small and rural providers GAO-23-105540. US Government Accountability Office. 2023 Apr 21. URL: https://www.gao.gov/products/gao-23-105540 [accessed 2025-02-19]
5. Baker C, Breidenstein R, Banuelos E. Road to a national health information network: US Trusted Exchange Framework and Common Agreement (TEFCA). Stud Health Technol Inform 2024 Jul 24;315:697-698. [doi: 10.3233/SHTI240285] [Medline: 39049387]
6. GNYHA lessons learned/preparation for future COVID-19 waves. Greater New York Hospital Association. URL: https://www.gnyha.org/wp-content/uploads/2020/08/LessonsLearned_SurgeSites.pdf [accessed 2025-02-19]
7. Mira JJ, Cobos A, Martínez García O, Bueno Domínguez MJ, Astier-Peña MP, Pérez Pérez P, et al. An acute stress scale for health care professionals caring for patients with COVID-19: validation study. JMIR Form Res 2021 Mar 09;5(3):e27107 [FREE Full text] [doi: 10.2196/27107] [Medline: 33687343]
8. Holmgren AJ, Apathy NC, Adler-Milstein J. Barriers to hospital electronic public health reporting and implications for the COVID-19 pandemic. J Am Med Inform Assoc 2020 Aug 01;27(8):1306-1309 [FREE Full text] [doi: 10.1093/jamia/ocaa112] [Medline: 32442266]

XSL•FO
RenderX

9. Guetterman T, Creswell JW, Kuckartz U. Using joint displays and MAXQDA software to represent the results of mixed methods research. In: McCrudden MT, Schraw G, Buckendahl C, editors. Use of Visual Displays in Research and Testing: Coding, Interpreting, and Reporting Data. Cambridge, MA: Information Age Publishing; 2015:145-175.
10. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013 Nov;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
11. Carayon P, Hundt AS, Karsh B, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec 01;15 Suppl 1(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
12. Kwan WM, Mok CK, Kwok YT, Ling H, Lam HW, Law TH, et al. Using the Systems Engineering Initiative for Patient Safety (SEIPS) model to describe the planning and management of the COVID-19 pandemic in Hong Kong. Int J Qual Health Care 2021 Mar 18;33(1):mzab042 [FREE Full text] [doi: 10.1093/intqhc/mzab042] [Medline: 33682006]
13. Carayon P, Smith MJ. Work organization and ergonomics. Appl Ergon 2000 Dec;31(6):649-662. [doi: 10.1016/s0003-6870(00)00040-5] [Medline: 11132049]
14. Cantor J, Qureshi N, Briscombe B, Chapman J, Whaley C. Association between COVID-19 relief funds and hospital characteristics in the US. JAMA Health Forum 2021 Oct;2(10):e213325 [FREE Full text] [doi: 10.1001/jamahealthforum.2021.3325] [Medline: 35977159]
15. Grogan CM, Lin YA, Gusmano MK. Health Equity and the Allocation of COVID-19 Provider Relief Funds. Am J Public Health 2021 Apr;111(4):628-631. [doi: 10.2105/AJPH.2020.306127] [Medline: 33539183]
16. Weiner-Lastinger LM, Pattabiraman V, Konnor RY, Patel PR, Wong E, Xu SY, et al. The impact of coronavirus disease 2019 (COVID-19) on healthcare-associated infections in 2020: a summary of data reported to the National Healthcare Safety Network. Infect Control Hosp Epidemiol 2022 Jan;43(1):12-25. [doi: 10.1017/ice.2021.362] [Medline: 34473013]
17. Ossom-Williamson P, Williams IM, Kim K, Kindratt TB. Reporting and availability of COVID-19 demographic data by US health departments (April to October 2020): observational study. JMIR Public Health Surveill 2021 Apr 06;7(4):e24288 [FREE Full text] [doi: 10.2196/24288] [Medline: 33821804]
18. Digital health transformation: what are maturity models. Healthcare Information and Management Systems. URL: https://www.himss.org/what-we-do-solutions/digital-health-transformation/maturity-models [accessed 2024-04-29]
19. Rokis K, Kirikova M. Exploring low-code development: a comprehensive literature review. Complex Syst Inform Model Quart 2023 Oct 31(36):68-86. [doi: 10.7250/csimq.2023-36.04]
20. The territorial impact of COVID-19: managing the crisis across levels of government. Organization for Economic Cooperation and Development. URL: https://tinyurl.com/ysjw5bp2 [accessed 2024-04-29]
21. Massoudi BL, Sobolevskaia D. Keep moving forward: health informatics and information management beyond the COVID-19 pandemic. Yearb Med Inform 2021 Aug 03;30(1):75-83 [FREE Full text] [doi: 10.1055/s-0041-1726499] [Medline: 34479380]
22. Maani N, Galea S. COVID-19 and underinvestment in the public health infrastructure of the United States. Milbank Q 2020 Jun 13;98(2):250-259 [FREE Full text] [doi: 10.1111/1468-0009.12463] [Medline: 32333418]
23. Kozlakidis Z, Abduljawad J, Al Khathaami AM, Schaper L, Stelling J. Global health and data-driven policies for emergency responses to infectious disease outbreaks. Lancet Global Health 2020 Nov;8(11):e1361-e1363. [doi: 10.1016/S2214-109x(20)30361-2]
24. Dron L, Dillman A, Zoratti MJ, Haggstrom J, Mills EJ, Park JJ. Clinical trial data sharing for COVID-19-related research. J Med Internet Res 2021 Mar 12;23(3):e26718 [FREE Full text] [doi: 10.2196/26718] [Medline: 33684053]
25. Aliabadi A, Sheikhtaheri A, Ansari H. Electronic health record-based disease surveillance systems: a systematic literature review on challenges and solutions. J Am Med Inform Assoc 2020 Dec 09;27(12):1977-1986 [FREE Full text] [doi: 10.1093/jamia/ocaa186] [Medline: 32929458]
26. The ONC cures act final rule: federal register. Office of the National Coordinator for Health Information Technology. URL: https://tinyurl.com/48dvbh53 [accessed 2024-04-29]
27. Electronic Case Reporting (eCR). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/ecr/php/index.html [accessed 2024-04-29]

Abbreviations

eCR: electronic case reporting
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
GNYHA: Greater New York Hospital Association
LNH: large network hospital
NYC: New York City
SEIPS: Systems Engineering Initiative for Patient Safety
SIH: small independent hospital

XSL•FO
RenderX

TEFCA: Trusted Exchange Framework and Common Agreement

Edited by E Mensah; submitted 26.06.24; peer-reviewed by M Gasmi , E Ford; comments to author 29.08.24; revised version received 29.10.24; accepted 10.11.24; published 26.03.25.

Please cite as:
Senathirajah Y, Kaufman DR, Cato K, Daniel P, Roblin P, Kushniruk A, Borycki EM, Feld E, Debi P
The Impact of the Burden of COVID-19 Regulatory Reporting in a Small Independent Hospital and a Large Network Hospital: Comparative Mixed Methods Study
Online J Public Health Inform 2025;17:e63681
URL: https://ojphi.jmir.org/2025/1/e63681
doi:10.2196/63681
PMID:40137048

©Yalini Senathirajah, David R Kaufman, Kenrick Cato, Pia Daniel, Patricia Roblin, Andre Kushniruk, Elizabeth M Borycki, Emanuel Feld, Poli Debi. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Use of Biometrics for Records Deduplication: Case Study of the National Data Repository in Nigeria

Ademola Oladipo[1*], MD, PhD; Ibrahim Dalhatu[2*], MD; Stephen Taiye Balogun[2], MD, MPH; Moyosola Bamidele[2], PhD; Ayodele Fagbemi[2], MPH; Isah Ahmed Abbas[2], BSc; Nannim Nalda[2], MPH; Richard Ugbena[2†], MD; Jude Orjih[2], MD; Timothy A Efuntoye[2], MD, MPH; Brooke Doman[1], MPH; Sadhna Patel[1], MPH; Herman Tolentino[1], MPH; Daniel Rosen[2], PhD; James Kariuki[1], MPH; Johnson Alonge[3], BTECH; Kehinde Balogun[3], MD, MPH; Nnamdi Umeh[3], MSc; Gibril Gomez[3], MSci; Oludare Onimode[3], MD, MPH; Olaposi Olatoregun[3], MD, MPH; Jay Osi Samuels[3], MD, MPH; Adebobola Bashorun[4], MD, MPH

[1]Division of Global HIV and TB, US Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA, United States

[2]Division of Global HIV and TB, US Centers for Disease Control and Prevention, Abuja, Nigeria

[3]Public Health Information, Surveillance Solutions and Systems Project, Abuja, Nigeria

[4]National AIDS and STI Control Program (NASCP), Ministry of Health, Abuja, Nigeria

[†]deceased

[*]these authors contributed equally

Corresponding Author:
Ademola Oladipo, MD, PhD
Division of Global HIV and TB, US Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA, United States

Abstract

Background: Nigeria has made significant investments in client-level electronic health systems, including the Nigeria Medical Record System (NMRS) and the National Data Repository (NDR), with funding from the US President's Emergency Plan for AIDS Relief through the US Centers for Disease Control and Prevention (US CDC). A biometric system was used across the US CDC–supported program in Nigeria to consistently track and monitor service uptake by people living with HIV during this period. The system was used to conduct deduplication analysis with the goal of preventing double counting and improving data integrity across all the US CDC-supported treatment sites (health facilities and community sites).

Objective: We describe the fingerprint biometric system in Nigeria and the process used for deduplicating health records of people living with HIV, including preliminary results.

Methods: The fingerprint biometric system leveraged the availability of the electronic NMRS at health facilities and the NDR. The integration of the fingerprint biometric module into the NMRS enabled fingerprints capture using SecuGen devices. Stakeholder engagement and capacity building were conducted with people living with HIV and health facility staff for fingerprint capture, storage, and transmission of the fingerprint templates to the NDR. Deduplication of the fingerprint templates was conducted in the automated biometric information system that is integrated with the NDR.

Results: We implemented fingerprint capture for 1,538,971 people living with HIV to deduplicate records from 1,141 treatment sites to improve the reliability and uniqueness of the system of records. Preliminary data showed that of the 1,538,971 records assessed by 30th June 2024, 1,520,187 of the active records (98.78%) had valid fingerprints, and 1,264,299 (83.17%) of the records with valid fingerprints were unique.

Conclusions: The implementation of a biometric system using fingerprint data allowed the identification of potentially duplicate records for resolution, thereby improving the quality of HIV treatment data for HIV program planning.

(Online J Public Health Inform 2025;17:e67580) doi:10.2196/67580

KEYWORDS

unique client identification; fingerprint; deduplication; data quality; PEPFAR-CDC; Nigeria

Introduction

Nigeria has made significant investments in client-level electronic health systems, including the Nigeria Medical Record System (NMRS) and the National Data Repository (NDR), with funding from the US President's Emergency Plan for AIDS Relief (PEPFAR) through the US Centers for Disease Control and Prevention (US CDC) [1]. The NDR is a centralized

XSL•FO
RenderX

repository of client-level data from disease programs of public health importance in Nigeria, starting with data from the HIV/AIDS program. The data from all HIV/AIDS-specific electronic medical records (EMRs) in Nigeria are transported and received securely using Secure Shell File Transfer Protocol, validated, and imported into the NDR. Figure 1 shows the high-level representation of the logical architecture of the NDR. The NDR data dictionary, information exchange standard, and case report message enable data exchange. The case-based message approach allows for the upload of only encounters that have not been pre-reported; thus, clients' records can be updated on the NDR and not replaced with each upload. Each message file submitted to the NDR has the data of a client that strictly follows the NDR information exchange standard. Data are XML-encoded in a hierarchical structure.

The NMRS is an electronic medical record system that was built on the OpenMRS software (Public Health Information, Surveillance Solutions and Systems Project, APIN Public Health Initiative), and it is used across 1141 health facilities and community sites in 19 states of Nigeria to manage records for people living with HIV supported by the PEPFAR program through the US CDC [1]. Community sites are unconventional locations outside of but linked to the health system, which are managed by community-based organizations (CBOs). The NMRS helps to improve the efficiency of collecting, managing, and reporting health information of people living with HIV [1-4]. Service delivery documentation takes place initially in the paper-based records; however, they are transferred by data entry clerks into the NMRS within 24 hours of documentation [1]. Records in the NMRS help health facilities to identify and verify clients when they receive services. The NMRS provides a comprehensive system for the availability and use of digital health records for decision-making.

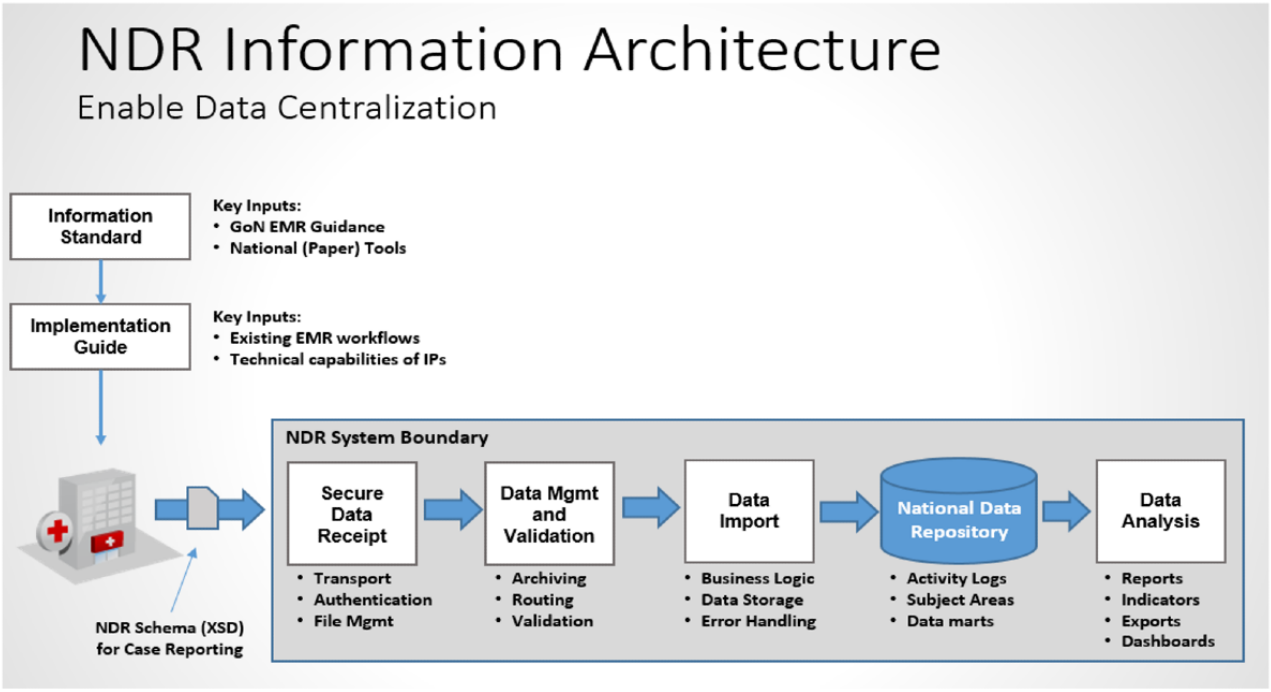
In US CDC-supported facilities, the NMRS was leveraged to implement the fingerprint capture initiative. The fingerprint is a widely acceptable form of biometric data that provides more definitive client identification [5-7]. Its use in identification is based on the science that no 2 individuals have the same fingerprint and that fingerprints remain consistent over a lifetime. The orientation of the ridges and valleys of each finger provides the distinction between any 2 individuals [6,7]. The uniqueness of fingerprints can be enhanced by the shape, number, and arrangement of the fingerprint patterns [6,7]. Stored fingerprint templates can enable client identification, verification, and authentication at the point of service, ensuring the provision of safe, secure, and effective care to the right client. This can improve the client treatment experience, as they can receive antiretroviral therapy services at any health facility of their choice and have their health records retrieved via the Patient Identification Management System [6-8]. Additionally, it assists in accounting for service linkage within and between sites and fewer medication errors due to misidentification.

People living with HIV are assigned hospital numbers and hand cards. These hand cards have composite identifiers, including hospital name, people living with HIV's name, hospital number, date of birth, and appointment schedule, which help to identify people living with HIV at the point of service delivery. However, the hand cards and hospital numbers are not universal across sites. As a result, when people living with HIV moved across sites or misplaced their hand cards, it was difficult to identify them, and new records might be created. Given this unique identification challenge, and to improve data quality and identification of unique records across all the 1,141 sites, the Government of Nigeria (GoN), in collaboration with the PEPFAR program, introduced fingerprint capture to be associated with individual medical records to address the problem. Stored fingerprint templates transported and received in the NDR are used for matching and deduplicating records to give account of the unique clients receiving treatment.

We present preliminary results and associated insights from the client records deduplication process we conducted using fingerprint data in the NDR.

XSL•FO
RenderX

Figure 1. The logical architecture of the National Data Repository (EMR: electronic medical record; GoN: Goverment of Nigeria; IP: implementing partners; Mgmt: management; NDR: National Data Repository; XSD: XML Schema Definition).



Methods

Ethical Considerations

This activity was reviewed by the US CDC, deemed not research, and was conducted consistent with applicable federal law and the US CDC policy. The requirements for written informed consent were waived as the preliminary analysis was summarized data by the NDR. Investigators did not interact with human subjects or access personally identifiable information.

Biometric Module in National Medical Records System

A biometrics module was integrated into the NMRS in 2019 to allow for fingerprint capture at the service delivery points. Fingerprints were captured with an optical technology, the SecuGen fingerprint scanner [9]. SecuGen has a range of features for individual users, software development kits (SDKs), and other related applications such as access control, time and attendance tracking, and identity verification [9,10]. The SecuGen Hamster Plus (Figure 2) was chosen as the preferred optical sensor due to its popularity, versatility, and advanced features, using the patented SEIR fingerprint biometric technology [9,11-14].

SecuGen offers a diverse range of fingerprint readers available in different forms, including desktop, handheld, and mobile devices [15]. These readers are designed to capture high-quality fingerprint images rapidly and accurately, making them ideal for deployment in high-traffic environments where speed and precision are crucial. The core technology driving SecuGen's capabilities lies in its proprietary SEIR fingerprint recognition algorithm, developed to provide precise and reliable fingerprint recognition in a wide array of scenarios [15]. This algorithm works in conjunction with high-quality fingerprint sensors to capture and analyze fingerprint data, comparing it against a database of previously registered fingerprints to authenticate users.

The biometric modules in both the enterprise and point of care NMRS were developed using SecuGen's SDK. This connects the SecuGen fingerprint reader to the computers and mobile phones. Fingerprints are stored in the NMRS while simultaneously linking them to the corresponding client records for which they were captured. By leveraging the functionality provided by the SecuGen SDK, the biometric module facilitated the seamless integration of fingerprint capture and storage within the NMRS.

XSL•FO
RenderX

Figure 2. SecuGen Hamster Plus fingerprint scanner.



Fingerprint Capture process

Fingerprint capture was integrated into the service delivery flow at the facility or community site. To ensure consistency of the fingerprint capture and reporting process, standard operating procedures were developed to guide health facilities, CBOs, and implementing partners to follow and conform to the recommended process.

To completely collect fingerprint data for all people living with HIV 15 years and older. Fingerprints for all new people living with HIV starting ART were captured during the initial enrollment at the medical records room as part of registration or triaging before clinical consultation. People living with HIV already receiving ART and returning to continue care were captured during medication refills or blood draw visits. Once valid fingerprints were collected for people living with HIV, they were not collected again at subsequent visits. This process was adapted for clients receiving treatment services in the community. The NMRS with the biometric module was installed on Android-based mobile phones. At the beginning of each week, community-based volunteers report to the health facility or CBO site they work for with the mobile phones to have the records of clients planned for services during the week preloaded. As services are provided in the field, fingerprints are captured for new and returning clients. At the end of the same week, the volunteers return to the site where records of services, including fingerprints collected, are synchronized with client records in the site's database.

To ensure comprehensive identification, we encouraged the capture of all fingerprints from all clients, with a minimum requirement of at least 6 fingers without repeating the capture of any finger. During the capturing process, the SecuGen device assessed the quality of each fingerprint captured. If the quality of a particular fingerprint was below 60%, the system flagged it for immediate repeat capture of that specific finger to ensure accurate and reliable data [11,16]. The biometric module was designed to indicate the quality of fingerprint capture. In addition, the biometric module had the capability to identify, flag, and disable fingerprint capture for clients that had valid fingerprints stored in the NMRS.

The captured fingerprints for each client were stored as templates and linked to the client's medical record and the unique identifier generated by the NMRS. As part of routine client record extraction and reporting, the captured fingerprint templates were automatically extracted along with the corresponding client record in an Extensible Markup Language file. This file was then reported to the NDR alongside other service delivery data. Within the NDR, the captured fingerprint templates were stored in the biometric table of the NDR database. The NDR implemented validation checks to ensure that all records in the biometric data table contained valid prints, that is, meeting the required minimum number of fingers and quality of capture.

NDR-ABIS Integration and Deduplication

An automated biometric identification system (ABIS) is a technological solution that uses biometric data to identify individuals [17-19]. ABIS MegaMatcher was designed to

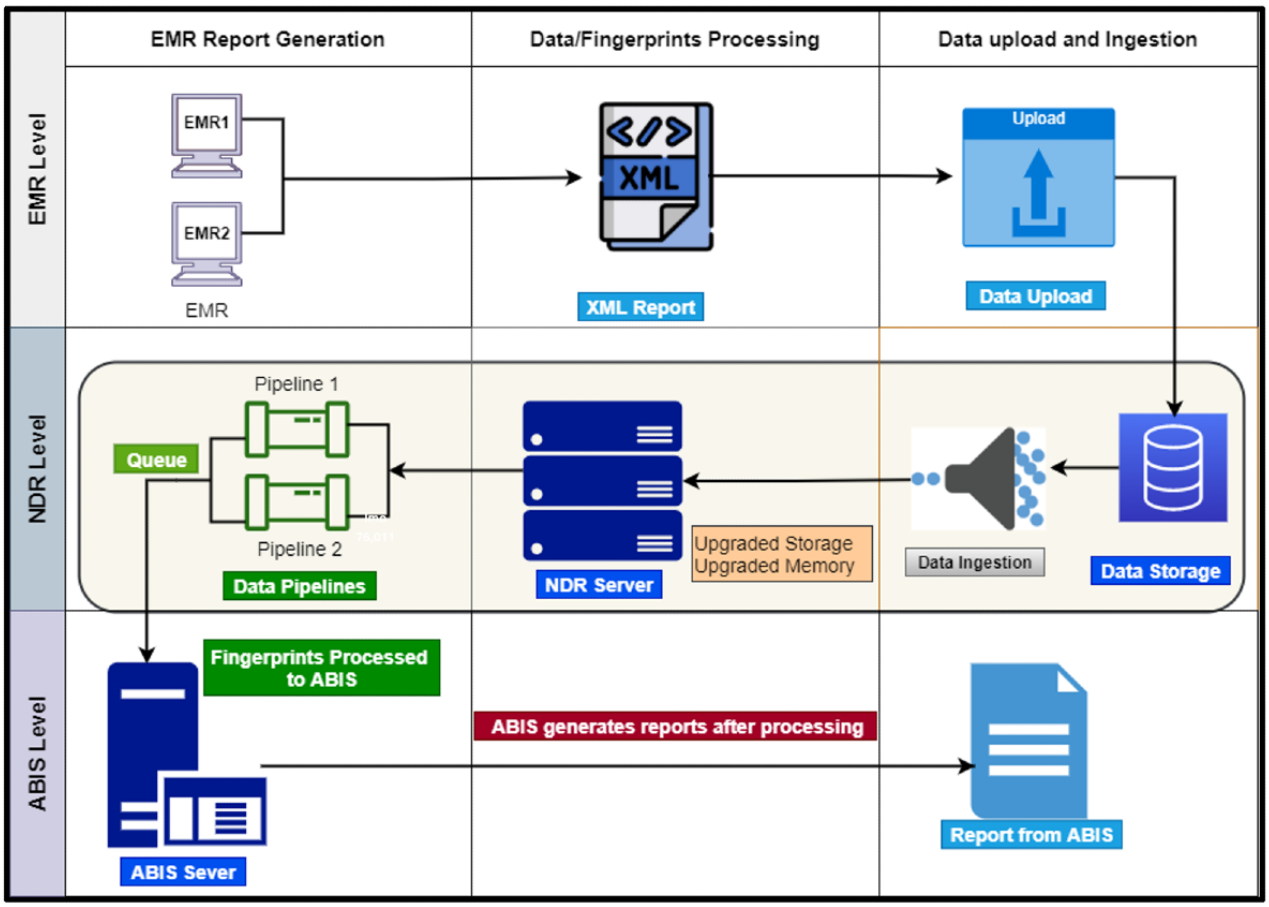
XSL•FO
RenderX

automate the process of matching and verifying biometric traits, such as fingerprints, against a database of pre-registered biometric records [18,19]. To enable fingerprint template matching and verification, the NDR established a bi-directional connection with the ABIS MegaMatcher. The NDR transmitted captured fingerprint templates to the ABIS MegaMatcher, which performed the matching process against other fingerprint templates in its database. The schematic diagram in Figure 3 illustrates an architectural diagram of the connection between the NMRS, NDR, and ABIS.

The ABIS MegaMatcher offers 2 deployment options: on-premises and cloud service [17]. In Nigeria, it was deployed as a cloud service to enhance accessibility.

The biometric matching process in the ABIS MegaMatcher involved comparing fingerprint templates from one client with fingerprints from all clients stored in the database, following a one-to-all matching approach for verification [17-19]. To ensure fast and accurate matching, the system used proprietary algorithms designed for high-performance biometric processing. These algorithms had the capability of performing up to 1.2 billion comparisons per second on each node of the system, thereby minimizing processing time and increasing accuracy. The outcome of matches in ABIS MegaMatcher is presented in Table 1.

Figure 3. A high-level fingerprint flow and connection between the Nigeria Medical Record System, National Data Repository (NDR), and automated biometric identification system (ABIS). EMR: electronic medical record.



XSL•FO
RenderX

Table . Outcome of matches in MegaMatcher ABIS (automated biometric identification system) using the Nigeria NDR (National Data Repository) fingerprint data as of 2023.

Outcome	Description
Enrolled	If the fingerprint sample does not match any enrolled templates in the database, the system declares the status as "Ok." This means that the provided fingerprint data are unique and do not correspond to any registered identity in the system.
Adjudication awaiting	In some cases, when performing a search across a large database, the system may return a list of potential matches or top candidates with a status of adjudication awaiting. These are individuals whose fingerprint data closely resemble the provided sample but may require further verification or human review to determine the exact identity.
Duplicate	When a fingerprint provided by an individual matches an enrolled template in the database, the system declares a match with the status duplicate found. This indicates that the individual's fingerprint data correspond to a previously registered identity.
Exception	This includes fingerprints affected by system configuration and calibration issues. Other reasons include technical errors or malfunctions such as system communication issues or hardware malfunctions during the transmission of the biometric status result.
No valid data	This is a status returned by NDR when the biometric data (fingerprint) does not adhere to NDR ISO standards or when there are incomplete or missing biometric features, or corrupt data from the electronic medical record.

Data Security

Rather than storing fingerprints as images, they were converted to templates and encrypted into Base64 hash before they were uploaded on the NDR. The fingerprint template is a condensed, more compact representation of a fingerprint, derived from the raw fingerprint vector data through feature extraction and template generation algorithms [20].

Encrypted fingerprint templates were pushed to ABIS, where they were decrypted and matched to identify duplicates. A role-based access control ensured that users without privileges cannot execute a particular operation.

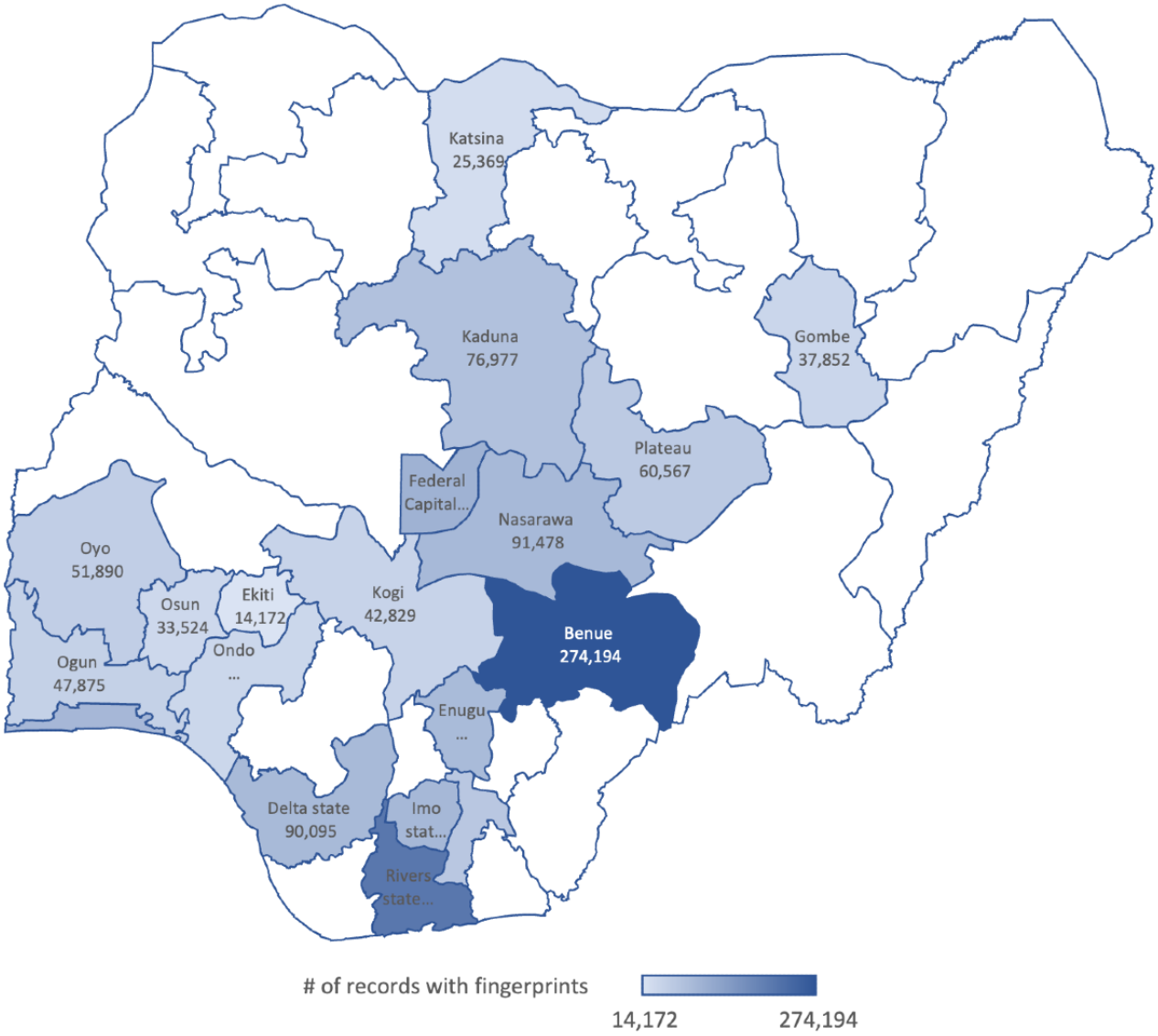
Duplicate fingerprints were matched by sex and date of birth outside of ABIS to reclassify duplicates as perfect and imperfect duplicates. Imperfect duplicates will be treated like the outcome, "Adjudication awaiting," which may require further verification or additional human review to arrive at a resolution.

Results

From 2021 through 2024, fingerprints were captured from 1,538,971 people living with HIV aged 15 years and older across 1141 health and community sites supported in 19 states (see Figure 4) by the US CDC in Nigeria. This represents 62% of the 2,483,916 fingerprint records captured across all 36 states and the Federal Capital Territory in Nigeria.

XSL•FO
RenderX

Figure 4. Distribution of fingerprints captured from people living with HIV in the Nigeria National Data Repository by State by June 2024 in CDC-supported states. CDC: Centers for Disease Control and Prevention.



Individual fingerprint templates were domiciled in the respective facility NMRS and reported to the NDR. Out of the 1,538,971 people living with HIV whose fingerprint records were collected, 1,520,187 (98.78%) had valid fingerprints (see Table 2). A valid fingerprint is a status returned by NDR when the biometric data (fingerprint) adhere to NDR ISO Standards and matching can be processed in the ABIS MegaMatcher.

Of the 1,520,187 valid fingerprint data assessed, 1,264,299 (83.17%) of the client records were authenticated as "enrolled," while 162,120 (10.66%) were determined as "duplicate," and 93,706 (6.16%) were awaiting adjudication, ie, the ABIS could not conclusively determine if these records should be authenticated as "enrolled" or if they were "duplicate" based on the fingerprint and would require further review. These are shown in Table 3.

Table 4 shows further review of duplicate fingerprints matched by sex and date of birth, and re-categorized into perfect duplicate, imperfect duplicate, and uncategorized (blanks). Perfect duplicates were records that matched when fingerprints, sex, and date of birth were compared; imperfect duplicates did not match with either sex or date of birth. Of the 162,120 duplicates found, 9591 (5.92%) were perfect duplicates, 18,854 (11.63%) were imperfect duplicates, and 133,675 (82.45%) were uncategorized (Null).

Table . Fingerprints processed in the ABIS (automated biometric identification system) MegaMatcher from the Nigeria National Data Repository data by June 2024.

Fingerprint status	Fingerprints (N=1,538,971), n	%
Valid data	1,520,187	98.78
No valid data	18,784	1.22

XSL•FO
RenderX

Table . Outcome of deduplication analysis from ABIS (automated biometric identification system) based on data from the Nigeria National Data Repository by June 2024 (for people living with HIV with valid fingerprints).

Fingerprint status	Client records (N=1,520,187), n	%
Adjudication waiting	93,706	6.16
Duplicate	162,120	10.66
Enrolled	1,264,299	83.17
Exception	62	0

Table . Manual matching of Nigeria National Data Repository fingerprints data by June 2024 with sex and date of birth to reclassify ABIS (automated biometric identification system) duplicate outcomes.

Duplicate type	Records (N=162,120), n	%
Perfect duplicate	9591	5.92
Imperfect duplicate	18,854	11.63
Null	133,675	82.45

Discussion

Principal Results

Our review observed 82.30% of the active client records with valid fingerprint capture were unique, highlighting the potential data quality gap that duplicate records pose to the integrity of a large HIV program. Integrating the fingerprint capture process into routine health care interactions and leveraging the capabilities of the NMRS, NDR, and ABIS MegaMatcher established a comprehensive and efficient system for capturing, storing, and establishing unique system records, thus verifying the uniqueness of clients' records in the NDR. This methodology helped to identify unique records of clients provided with services, thus preventing double counting. Fingerprint implementation helped to improve the quality of data on the NDR and will help to improve the utility of data from the NDR for programmatic reporting, case-based surveillance, and client-centric monitoring. The number of unique clients receiving treatment is a critical input into the UNAIDS SPECTRUM modeling tool, and better reporting will improve estimation of people living with HIV burden in Nigeria.

A clear resolution protocol that highlights the feedback communication of specific duplicate type or invalid fingerprint records to the health facility or CBO for immediate investigation and resolution is needed to address the different types and patterns of duplicates across locations. Clients with duplicate or invalid fingerprints should be engaged before final determination of validity is reached.

Good-quality data are fundamental to planning public health interventions. Clients may, for reasons of seeking livelihood, migrate from place to place during lifelong HIV treatment without following the standard transfer protocols [21]. These clients could be received as new clients, given new client identification numbers, and reported to the NDR as separate individuals. A pilot study from Malawi was able to determine that approximately 8% of people living with HIV in the study population had more than one client ID in EMR [22]. Fragmented health care systems, client mobility, and inefficient care coordination are some reasons for multiple client registration [23]. Multiple registrations increase the chance of counting clients more than once, thus providing inaccurate reported results. Accurate data are critical for appropriate program planning activities like target setting and budgeting for resources in public health programs.

Fingerprint biometrics can assist with tracking services received by clients, which makes it a useful identifier for longitudinal client monitoring and case-based surveillance [24,25]. Similarly, WHO suggested the use of client biometrics, including fingerprints, for person-centered HIV monitoring and case surveillance [26]. In addition to deduplication of clients using fingerprint records in the NDR for accurate reporting of the number of clients receiving treatment, fingerprints can be used to track sentinel events among HIV cases to inform clinical, programmatic, and public health action.

Public Health Implications

Biometrics such as fingerprints and the ability to use them to deduplicate clients are critical to program monitoring and planning, effective service linkage, client safety, and person-centered clinical management.

Programs set operational targets each fiscal year. Setting targets depends on factors including target benchmark, resources, and the epidemic gap to be closed. The use of client fingerprints to identify and remove duplicate records is necessary for accurately setting targets and avoiding wastage of resources. Targets help determine resource allocation and budgeting, amplifying the need for good quality data.

Focusing on the scope of sustainability, it will be crucial for countries to have accurate client-level data so they are able to efficiently use funds, make projections, conduct HIV case-based surveillance, and move resources to where they are needed based on the unmet population needs.

Conclusions

Large-scale deduplication like we have presented was possible because of the NDR, the centralized data warehouse. Integrating fingerprint biometrics into routine HIV program implementation can potentially identify unique client records, prevent double

XSL•FO
RenderX

counting, and improving the reliability of records as belonging to unique individuals. The results highlight the potential data quality gap posed by duplicate records, emphasizing the importance of accurate data for programmatic reporting and client-centric monitoring.

Limitations

The quality of the fingerprint data analysis is dependent on several factors, namely, user, technological (SecuGen, extraction and upload to the NDR, and storage), and the ABIS analytic engine. This paper has not described the methods and algorithms for deduplication of fingerprints in ABIS because ABIS is proprietary.

Acknowledgments

We acknowledge the funding support of the President's Emergency Plan for AIDS Relief (PEPFAR) through the US Centers for Disease Control and Prevention (US CDC; grant number GH002346), the leadership role of the Government of Nigeria for successful implementation, the implementing partners for facilitating reporting, the health care workers for service delivery, and the program beneficiaries for availing themselves of the services. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the funding agencies.

Conflicts of Interest

None declared.

References

1. Dalhatu I, Aniekwe C, Bashorun A, et al. From paper files to web-based application for data-driven monitoring of HIV Programs: Nigeria's journey to a national data repository for decision-making and patient care files to web-based application for data-driven monitoring. Methods Inf Med 2023 Sep;62(3-04):130-139. [doi: 10.1055/s-0043-1768711] [Medline: 37247622]
2. Boyd AT, Jahun I, Dirlikov E, et al. Expanding access to HIV services during the COVID-19 pandemic-Nigeria, 2020. AIDS Res Ther 2021 Sep 19;18(1):62. [doi: 10.1186/s12981-021-00385-5] [Medline: 34538268]
3. Boyd AT, Ogbanufe O, Onyenuobi C, et al. Scale-up of antiretroviral treatment access among people living with HIV in Rivers State, Nigeria, 2019--2020. AIDS 2021 Jun 1;35(7):1127-1134. [doi: 10.1097/QAD.0000000000002858] [Medline: 33946087]
4. Dirlikov E, Jahun I, Odafe SF, et al. Rapid scale-up of an antiretroviral therapy program before and during the COVID-19 pandemic - Nine States, Nigeria, March 31, 2019-September 30, 2020. MMWR Morb Mortal Wkly Rep 2021 Mar 26;70(12):421-426. [doi: 10.15585/mmwr.mm7012a3] [Medline: 33764965]
5. Etter L, Simukanga A, Qin W, et al. Project SEARCH (Scanning EARs for Child Health): validating an ear biometric tool for patient identification in Zambia. Gates Open Res 2020;4:168. [doi: 10.12688/gatesopenres.13197.1] [Medline: 33655198]
6. Fatima K, Nawaz S, Mehrban S. Biometric authentication in health care sector: a survey. 2019 Presented at: p. 1-10. [doi: 10.1109/ICIC48496.2019.8966699]
7. Mason J, Dave R, Chatterjee P, Graham-Allen I, Esterline A, Roy K. An investigation of biometric authentication in the healthcare environment. Array 2020 Dec;8:100042. [doi: 10.1016/j.array.2020.100042]
8. Etter LP, Ragan EJ, Campion R, Martinez D, Gill CJ. Ear biometrics for patient identification in global health: a field study to test the effectiveness of an image stabilization device in improving identification accuracy. BMC Med Inform Decis Mak 2019 Jun 18;19(1):114. [doi: 10.1186/s12911-019-0833-9] [Medline: 31215427]
9. Hamster plus fingerprint reader. SecuGen. 2025. URL: https://secugen.com/products/hamster-plus/ [accessed 2023-06-07]
10. Software developer kits. SecuGen. 2025. URL: https://secugen.com/products/sdk/ [accessed 2023-06-07]
11. Aseri V, Nagar V, Godara V, et al. A comparative study on scanned fingerprint after applying lubricants and without scanned fingerprint on porous surface (white paper). 2022 Presented at: Mater Today Proc p. 1515-1531. [doi: 10.1016/j.matpr.2022.04.007]
12. Azeta AA, Iboroma DOA, Azeta VI, Igbekele EO. Implementing a medical record system with biometrics authentication in e-health. In: 2017 IEEE AFRICON 2017:979-983. [doi: 10.1109/AFRCON.2017.8095615]
13. Yahaya MM, e-Extension Department National Agricultural Extension and Research Liaison Services, Ahmadu Bello University Zaria, Nigeria. Fingerprint based attendance management system. IJARCS 2022;13(1):7-10. [doi: 10.26483/ijarcs.v13i1.6802]
14. Ogundepo OY, Omeiza IOA, Akpojaro MA. Development of a real time fingerprint authentication/identification system for students' record. NIJOTECH 2019;38(1):212. [doi: 10.4314/njt.v38i1.26]
15. Products: fingerprint readers. SecuGen. 2025. URL: https://secugen.com/products/ [accessed 2023-06-07]
16. Abhyankar A, Kulkarni N, Kumar S, Stephanie S. Fingerprint image quality and prediction of matching performance. CiiT Int J Digit Image Process ;1:45-49 [FREE Full text]

XSL•FO
RenderX

17. MegaMatcher automated biometric identification system. Neurotechnology.: Laisves; 2021. URL: https://www.neurotechnology.com/megamatcher-abis.html [accessed 2024-08-25]
18. Anufriiev P, Bashkov Y, Khoma D. Experimental Face Recognition System Based On Improved Artificial Intelligence Model 2022:181-186.
19. Nlend JAB, Nguena IB, Bouetou TB. Large scale deduplication based on fingerprints. arXiv. Preprint posted online on Jan 13, 2021. [doi: 10.48550/arXiv.2101.04976]
20. Jain AK, Flynn P, Ross AA. Hanbook of Biometrics: Springer; 2007.
21. Etoori D, Kabudula CW, Wringe A, et al. Investigating clinic transfers among HIV patients considered lost to follow-up to improve understanding of the HIV care cascade: findings from a cohort study in rural north-eastern South Africa. PLOS Glob Public Health 2022;2(5):e0000296. [doi: 10.1371/journal.pgph.0000296] [Medline: 36962304]
22. Bengtson AM, Kumwenda W, Lurie M, et al. Improving monitoring of engagement in HIV care for women in option B+: a pilot test of biometric fingerprint scanning in Lilongwe, Malawi. AIDS Behav 2020 Feb;24(2):551-559. [doi: 10.1007/s10461-019-02748-6] [Medline: 31773445]
23. Retention in HIV programmes: defining the challenges and identifying solutions: meeting report (13-15 September 2011, Geneva, Switzerland). World Health Organization. 2012. URL: https://www.who.int/publications/i/item/9789241503686 [accessed 2025-07-30]
24. Kavanagh MM, Baral SD, Milanga M, Sugarman J. Biometrics and public health surveillance in criminalised and key populations: policy, ethics, and human rights considerations. Lancet HIV 2018 Oct 5;doi:30243-30241. [doi: 10.1016/S2352-3018(18)30243-1] [Medline: 30305236]
25. Serwaa-Bonsu A, Herbst AJ, Reniers G, et al. First experiences in the implementation of biometric technology to link data from Health and Demographic Surveillance Systems with health facility data. Glob Health Action 2010 Feb 24;3. [doi: 10.3402/gha.v3i0.2120] [Medline: 20200659]
26. Consolidated Guidelines on Person-Centered HIV Patient Monitoring and Case Surveillance: WHO; 2017.

Abbreviations

ABIS: automated biometric identification system
CBO: community-based organization
EMR: electronic medical record
NDR: National Data Repository
NMRS: Nigeria Medical Record System
PEPFAR: US President's Emergency Plan for AIDS Relief
SDK: software development kit
US CDC: US Centers for Disease Control and Prevention

Edited by E Mensah; submitted 18.10.24; peer-reviewed by C Aniekwe, EH Lee, O Adedokun; revised version received 07.05.25; accepted 20.06.25; published 26.08.25.

Please cite as:
Oladipo A, Dalhatu I, Balogun ST, Bamidele M, Fagbemi A, Abbas IA, Nalda N, Ugbena R, Orjih J, Efuntoye TA, Doman B, Patel S, Tolentino H, Rosen D, Kariuki J, Alonge J, Balogun K, Umeh N, Gomez G, Onimode O, Olatoregun O, Samuels JO, Bashorun A
Use of Biometrics for Records Deduplication: Case Study of the National Data Repository in Nigeria
Online J Public Health Inform 2025;17:e67580
URL: https://ojphi.jmir.org/2025/1/e67580
doi:10.2196/67580

© Ademola Oladipo, Ibrahim Dalhatu, Stephen Taiye Balogun, Moyosola Bamidele, Ayodele Fagbemi, Isah Ahmed Abbas, Nannim Nalda, Richard Ugbena, Jude Orjih, Timothy A Efuntoye, Brooke Doman, Sadhna Patel, Herman Tolentino, Daniel Rosen, James Kariuki, Johnson Alonge, Kehinde Balogun, Nnamdi Umeh, Gibril Gomez, Oludare Onimode, Olaposi Olatoregun, Jay Osi Samuels, Adebobola Bashorun. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 26.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Open Source Web Application (HealthTest) for Emotional Health and Wellness Management in University Students: Development and Usability Study

Lucrecia Llerena[*], PhD; Daisy Nata Castro[*], MSc; Nancy Rodriguez[*], MSc; Donald Silva Sánchez[*], BEng

Carrera de Ingeniería de Software, Facultad de Ciencia de la Computación y Diseño Digital, Universidad Técnica Estatal de Quevedo, Vía Quevedo-Santo Domingo, Km.1 1/2, Quevedo, Ecuador

[*]all authors contributed equally

Corresponding Author:
Lucrecia Llerena, PhD
Carrera de Ingeniería de Software, Facultad de Ciencia de la Computación y Diseño Digital, Universidad Técnica Estatal de Quevedo, Vía Quevedo-Santo Domingo, Km.1 1/2, Quevedo, Ecuador

Abstract

Background: Emotional health plays a fundamental role in quality of life, particularly after the COVID-19 pandemic, which has increased stress and anxiety, especially among children and young people.

Objective: This study aimed to focus on the early identification of emotional processes that affect individuals' well-being and their effective management.

Methods: The open-source web app HealthTest was developed to help users understand and manage their emotions through tests focused on aspects such as stress, anxiety, and depression. The Open Source Scrum (OSCRUM) framework was used to optimize collaboration and effectively achieve objectives.

Results: HealthTest has established itself as a valuable tool for mental health professionals by gathering data from seventh-semester software engineering students and external users. It identifies trends in stress, anxiety, and depression through user self-assessments. In addition, it provides meditation and relaxation resources designed to support users in managing their emotional well-being.

Conclusions: This study promotes accessibility to self-care and health care tools. HealthTest reaffirms its commitment to benefiting both mental health professionals and patients, providing an effective avenue for controlling and improving emotional well-being.

(Online J Public Health Inform 2025;17:e69413) doi:10.2196/69413

KEYWORDS

emotional health; open source; stress; anxiety; well-being of the person

Introduction

Background

Currently, society is in the process of recovering from the COVID-19 pandemic, which significantly disrupted mental health services worldwide, causing widespread interruptions in psychological counseling and psychotherapy [1,2]. A 2020 study conducted in Spain revealed that female health care professionals experienced higher levels of anxiety, stress, and depression compared to their male counterparts, with these issues being more prevalent among individuals aged 20 to 37 years [3].

The prevention and treatment of mental health issues are crucial due to their impact on quality of life and daily functioning [4,5]. With technological advancements, web applications have emerged as valuable tools across various sectors, including health care [6]. Maintaining optimal mental health enables individuals to face challenges effectively, boost productivity, and enhance interpersonal relationships. Extensive research through a Systematic Mapping Study (SMS) has identified key methodologies, techniques, and tools to address these mental health issues, highlighting their importance and applicability in practical contexts [7].

Following the research phase, the Open Source Scrum (OSCRUM) framework was selected for the development of the web application. OSCRUM is an adaptation of the scrum model for open-source software, introduced in this study. This choice is based on OSCRUM's effectiveness in providing an organized and structured method that facilitates the planning, tracking, and evaluation of objectives in the health care sector [8].

The COVID-19 pandemic has had emotional and social impacts, particularly on children and adolescents [9]. This highlights how the lack of emotional well-being assessment affects

XSL•FO
RenderX

academic performance, interfering with the ability to process information and learn, ultimately deteriorating academic outcomes [10,11].

In response to this need, we have developed HealthTest, an open-source web application built using the OSCRUM framework. Its goal is to provide a platform for assessing anxiety, stress, and depression. HealthTest allows the collection of both anonymous and nonanonymous data on users' emotional well-being, whether they are registered on the platform or not. It delivers clear results, explanations, and recommendations based on reliable psychology sources.

HealthTest distinguishes itself through its transparency and collaborative approach, offering health professionals an intuitive and visually engaging interface to assess and monitor patients' emotional well-being. As development progresses, it is envisioned as a dynamic tool for collecting emotional data, facilitating more informed and personalized care. To achieve this, the OSCRUM framework was implemented, enhancing teamwork and goal attainment through structured planning, development, review, and approval processes. OSCRUM's adaptability to workflow dynamics ensures agile and flexible solutions, allowing HealthTest to evolve in response to client needs and shifting work environments.

This study aims to assess the usability and impact of the HealthTest app in delivering accessible tools for emotional health management, emphasizing its open-source nature and collaborative development under the OSCRUM framework. Beyond detailing the development of HealthTest, it also introduces OSCRUM as an innovative approach to open-source digital health solutions. Unlike traditional methodologies, OSCRUM facilitates continuous, community-driven updates, positioning HealthTest as a dynamic, evolving tool rather than a static intervention.

Related Works

The systematic search focused on identifying open-source web applications developed in the field of emotional health. As a result of this search, several relevant studies were identified, which had a positive impact on the development of the new web app [12,13]. In the field of emotional and mental health, several open-source tools have been developed to enhance assessment and diagnosis. For instance, MindLogger, a platform that facilitates the creation of clinical assessments and the scheduling of customized questionnaires for mental health and learning disorders [12]. In addition, drug-use insights, a web-based tool designed to analyze the specific and esoteric language used on social media related to substance abuse [14].

In the management and analysis of medical and genomic data, innovations such as PLCOjs stand out [15]. This SDK enables the creation of interactive graphical representations and analytical workflows for genomic data. Similarly, introduced Leaf, a web application that provides an intuitive interface for querying clinical databases, offering more flexible and direct access to medical information [16].

The development of personalized clinical trials and collaborative tools has also advanced significantly. StudyU, an open-source platform, facilitates the execution of N-of-1 trials [17]. In addition, an international pilot project explored the use of open-source Enterprise Resource Planning (ERP) for teaching hospital information system development, fostering collaboration and enhancing health care infrastructure in developing countries [18].

In the pediatric care sector, an innovative collaborative framework based on ARCore has been introduced to enhance children's experiences during medical procedures through augmented reality applications. This approach fosters more child-friendly and less stressful interactions for young patients [19].

Finally, enhancing the understanding and management of health information has been the focus of several significant studies. One study explored the automatic identification of sections in Electronic Clinical Narratives, improving the processing and analysis of medical records in Spanish [20]. Meanwhile, other research highlights the importance of adopting dynamic, open-source web technologies to foster collaboration and information exchange in health informatics [13].

After extensive analysis of various studies, it can be concluded that numerous research efforts in the health care sector have focused on the development of web apps. However, there is a noticeable gap in research addressing specific areas, such as mental and emotional health. Most attention has been directed toward other fields, leaving a void in exploring how web apps could contribute to these critical aspects.

Methods

Ethical Considerations

This study was approved by the Research and Ethics Committee of the Universidad Técnica Estatal de Quevedo on June 27, 2025 (Certificate No. CERT-ÉTICA-013-DICYT-2025). The committee determined that the research complies with national and international ethical standards and does not involve experimentation on living human beings, in accordance with the Ecuadorian CEISH regulations (Registro Oficial No. 279, July 1, 2014). All participants—adult university students and collaborators—voluntarily agreed to participate and provided signed informed consent. The consent form included all required information regarding the purpose, procedures, and data protection measures of the study. All data were anonymized and stored securely to ensure participant confidentiality. No financial or material compensation was provided to participants.

Definition of Development Activities in OSCRUM

OSCRUM is an open-source framework that adapts the agile SCRUM model [8], focusing on collaboration, transparency, and collective decision-making in software development. It incorporates strategic planning, project management, documentation, bug tracking, and testing to ensure software quality. Open-source software development using OSCRUM is crucial as it fosters collaborative participation and essential contributions. This approach enables early detection and resolution of issues while ensuring that the software meets quality standards through a meticulous testing strategy.

XSL•FO
RenderX

Within this framework of OSCRUM, several key players emerge, including (1) the Lead Maintainer, (2) the Core Contributor, (3) the Community, (4) Sprint, and (5) the Backlog. These roles collaborate harmoniously to ensure the project's objectives are achieved. Ultimately, OSCRUM provides a robust framework for creating open-source software, fostering efficiency and collaboration throughout the development process. SCRUM and its open-source adaptation, OSCRUM, share characteristics such as continuous delivery, team collaboration, structured cycles, and a customer-centric approach. Rahman et al [8] highlight that OSCRUM is structured around eleven defined activities, emphasizing regular meetings and clear processes.

Involves Identifying Problems and Recruiting Volunteers

Through brainstorming and competitive analysis to define the initial requirements for HealthTest, a blog was used for the community to contribute ideas and desired features for the web system.

Focuses on Communication With Collaborators

After gathering information and requirements through observation and document analysis, a meeting with the development team is planned to define the functional and nonfunctional requirements of the web app. These will be published in the web artifact for users to track progress.

Initial Launch Planning

Involves a meeting with the development team to organize sprints for the development of the web application under the OSCRUM framework. An interface network system, role identification, and the creation of the first sprint were established, ensuring that each work cycle maintains an adequate pace within the planned timeline.

Launch Plan and Product Status

After completing the sprint backlog of features, activities are listed in the launch plan with their respective status, and the list of features is published in the web artifact.

Functionality Updates

Sprints are updated with input from community users through selected ideas to improve the functionality of the web application.

Testing the Source Code

A GitHub repository is created for the HealthTest code, facilitating collaboration and code access. Subsequently, the maintainer and key contributors review the features during a sprint retrospective meeting and publish the project on the web, accompanied by documentation and screenshots. To collect bug reports, a blog explaining the system's functionality is used, and user participation is encouraged through promotion on social media and online thematic groups. This ensures active feedback and continuous project improvement.

Reporting Bugs

In OSCRUM, the community plays a key role in ensuring the project meets user needs. During sprint reviews, members have the opportunity to evaluate completed features and report issues or bugs. This process is essential for optimal project development and achieving proposed objectives. Community feedback is vital for improving project quality and ensuring its future success.

External Contributor Entry

OSCRUM, focused on collaboration and teamwork for open-source software development, emphasizes incorporating external contributors. These contributors are selected by the main maintainer, and their participation is crucial for the project's efficient execution. They bring valuable expertise and knowledge, increasing community engagement and software quality. In addition, a bug reporting template is provided [21], promoting effective collaboration and continuous development.

Bug Fixing

Collaborators address and resolve issues encountered during development, enhancing software quality and meeting user expectations. This preventive action ensures the project's success and avoids future problems.

Solution Approval

After uploading the project to the repository with the corrected features, the community will review the web app once again.

The implementation of HealthTest followed a structured approach based on OSCRUM principles, ensuring a systematic and iterative development process. The key activities involved are summarized in Table 1.

XSL•FO
RenderX

Table . OSCRUM activities for web application development.

N	Criterion	Definition
1	Discovery of the problem and search for volunteers	An individual or a small group brainstorms to discover the problem
2	Communication	Meetings are held with the developers to obtain the software functionalities
3	Initial launch planning meeting	Meetings are held to create sprints and carry out the first functionalities
4	Release plan and status	Once the main functionalities are finished, a test launch is carried out
5	Feature update	After the test release the features of the web application are updated
6	Test source code	The community can test the source code of the web application
7	Bug report	When the community tests the web application, it will file a bug report for the developers to correct
8	Contributions from external members	External people are expected to contribute to updates and new features of the web application
9	Repair	Once the error report is obtained by the community, the repair is carried out so that the software does not have any errors
10	Approval (Validation)	A new evaluation is carried out to find out if the web application has any errors and if it does not, the web application is approved

Implementation of the Health Test Web Application in OSCRUM

The development of the HealthTest web application will be conducted within the OSCRUM framework, ensuring an agile and structured open-source development process. The following steps will be undertaken:

Frontend Development

The web application's interface will be designed using the Angular framework, ensuring responsiveness, accessibility, and usability across various devices and user types. The development team will implement an intuitive navigation structure, enabling users to access key functionalities with efficiency and ease.

Backend Development

The backend will be developed using Spring, a Java-based framework that facilitates secure data processing, API communication, and the implementation of business logic. RESTful services will be designed to efficiently manage data interactions between the frontend and backend.

Database Management

PostgreSQL will be used to store and manage user data, ensuring structured and optimized queries for enhanced performance. The database relationships will be designed to maintain data integrity and enable seamless retrieval of stored information [22,23].

Results

This section presents the results obtained from applying the OSCRUM framework [8] focusing on the implementation of the HealthTest web application and access to its source code.

Application of the OSCRUM Framework in HealthTest Development

This section details each activity of the OSCRUM framework used for the development of the HealthTest web application:

Problem Discovery and Volunteer Recruitment

Most web applications for emotional well-being are paid services. HealthTest was created as an open-source web application offering tests for stress, anxiety, and depression. After completing these tests, HealthTest provides recommendations and solutions, including specialized blogs and videos, to help users overcome or prevent these issues.

A web artifact was developed to gather feedback on a web application, engaging regular users of open-source tools. In this space, collaborators conducted brainstorming sessions and a competitive analysis to define the team's requirements, fostering active community participation in the project's development [24].

Communication

This section outlines the functional and nonfunctional requirements for HealthTest, derived from techniques such as brainstorming and web artifacts (Blog), which streamlined the information-gathering process. The results and the general use case diagram, illustrating client interactions with the HealthTest

XSL•FO
RenderX

web application, are available in the supplementary materials [25,26].

Initial Launch Planning Meeting

After completing the initial activities, interfaces, roles, and high-fidelity prototypes for HealthTest were defined using Angular Command Line Interface. (Figure 1) presents the prototype of the main page. Planning meetings established agreements on task development and assigned timelines. The management of this planning phase is conducted through the repository hosting the HealthTest source code.

Figure 1. Main interface of HealthTest.



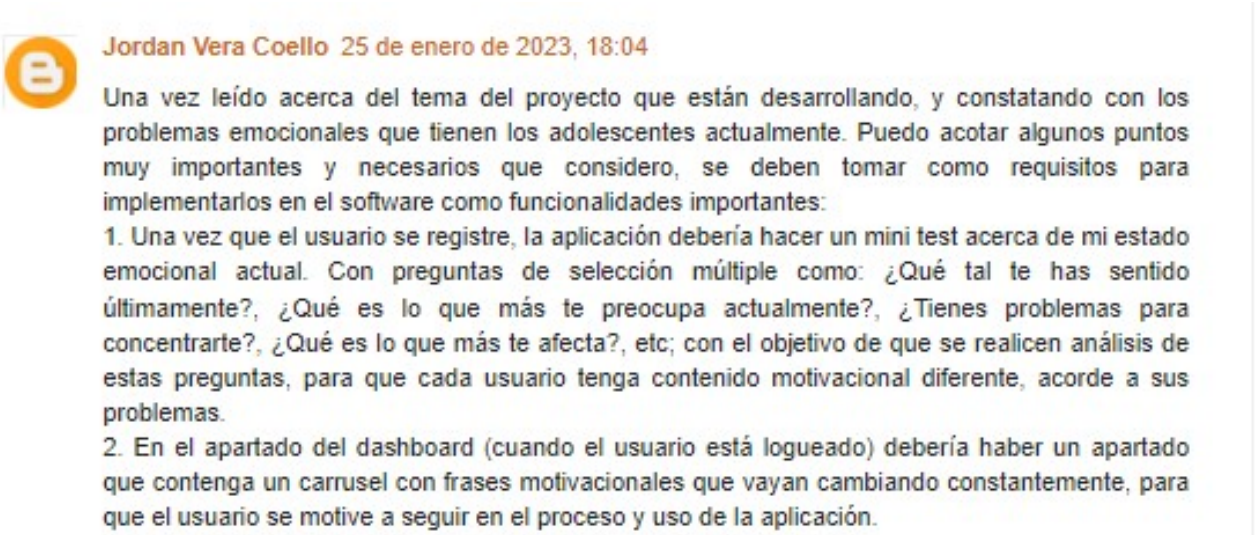
Launch Plan and Status

The launch planning for HealthTest ensures an on-time release that meets community expectations. The team collaborated with users to adapt the product to their needs, complemented by interface reviews to ensure functionality.

Functionality Improvement

The functionalities of the HealthTest web application were updated based on feedback from a specialized blog and pilot tests with diverse users, including seventh-semester software engineering students from the Technical State University of Quevedo. These students were seven, selected for their technological expertise, tested HealthTest in various scenarios. The results helped the development team identify and resolve issues, enhancing the application's functionality and quality. The students' valuable feedback significantly contributed to the project's success, enabling HealthTest to meet the expectations of a broader audience. Figure 2 displays a snippet of the feedback published by users on the blog web artifact [24].

Figure 2. User feedback.



XSL•FO
RenderX

Source Code Testing

Tests were conducted on the HealthTest web application with 7 users to identify errors. Friends and colleagues of the researchers were recruited through invitations to participate in the HealthTest blog, where they expressed interest and provided their email addresses for future contact. Using these emails, instructions and resources were sent to test the source code and report errors. HealthTest was hosted on the free Netlify server, where users followed specific instructions to evaluate the application and contribute to diagnosing and resolving identified issues.

Bug Reporting

A process was established to identify bugs in HealthTest, recruiting experts to evaluate the web application. These experts were provided with a specific template to report bugs, allowing them to identify issues and generate detailed reports. This practice is crucial to ensuring software quality and enhancing the user experience. Thanks to the feedback, the development team quickly corrected the errors and improved the application's functionality. The bug reports were integrated into the sprints for resolution, and the complete reports are available in the supplementary materials [27].

Contributions From External Members to the Project

The HealthTest web application has its code publicly available on GitHub [28]. Anyone can review the code, but write permissions are required to collaborate, which are granted by the project's main maintainer. This maintainer oversees and approves external contributions, ensuring the quality and consistency of the code. The use of GitHub as a hosting platform, along with permission controls, guarantees the transparency and security of HealthTest's source code.

External users from the open-source community provided several key suggestions after reporting bugs in HealthTest, including errors during registration, inadequate handling of email input and user type during registration, and static screens for the "Administrator," "Psychologist," and "Patient" roles. In addition, they recommended that the login screen display the role and name of the user logging in and noted that the logout button was nonfunctional. These insights were crucial for identifying and addressing issues to improve the application's functionality.

Repair

Since issues were identified in previous stages, appropriate solutions were developed, and HealthTest was updated based on the feedback provided by external collaborators to meet user needs.

Validation

After addressing the errors identified in HealthTest during the evaluation with external users, a new review is conducted to verify that the issues with the functionalities have been resolved. Once the corrections are confirmed, the web application is approved. The implemented corrections include: (1) a registration form was implemented for all user types, including patients, psychologists, and administrators, simplifying the registration process; (2) each input was validated, ensuring the accuracy and security of the provided information; and (3) a problem with the password recovery system was resolved, which previously failed to send reset emails effectively.

The HealthTest development team evaluated and approved each of the corrections made to the web application, confirming that these solutions were effective. Thanks to these improvements, users now enjoy a smoother and more secure experience when using the application.

Implementation and Deployment of HealthTest

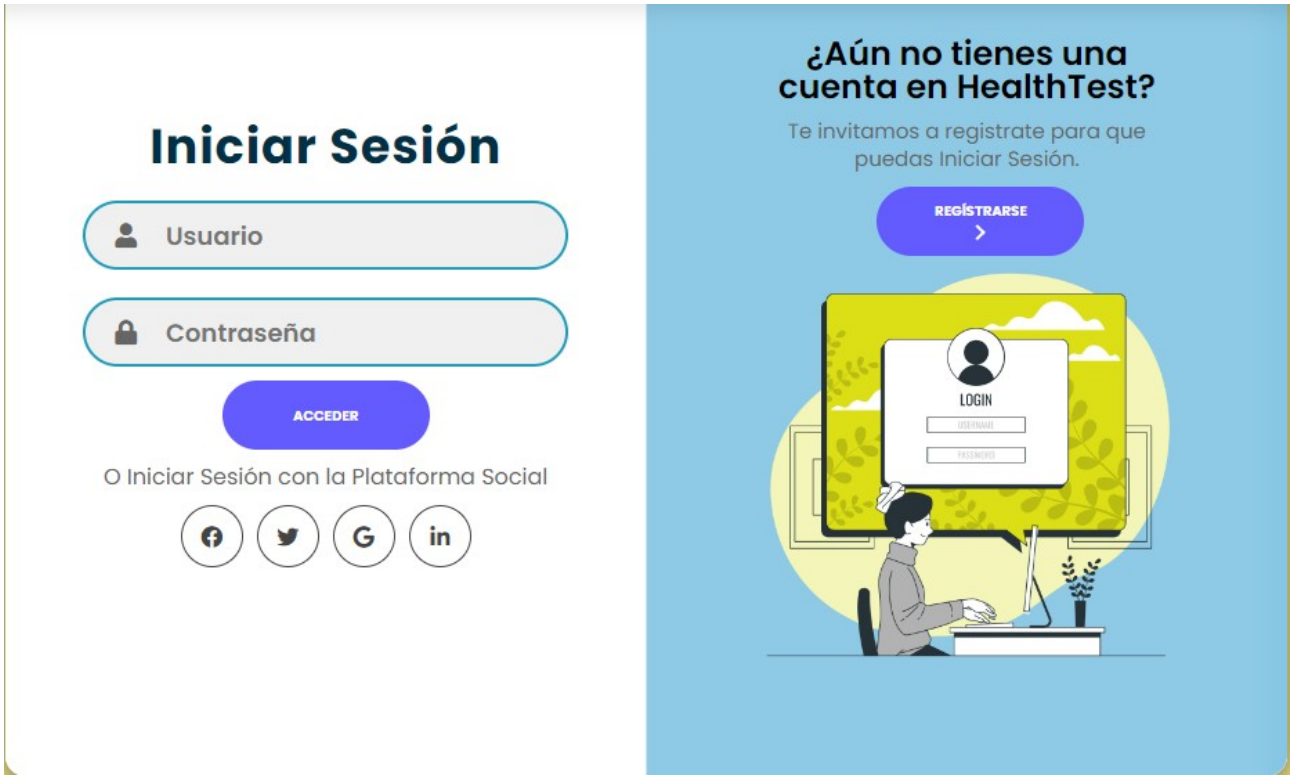
The HealthTest web application was successfully developed and implemented following the OSCRUM framework, meeting the planned objectives for each key component:

Frontend Development

The final user interface was built using Angular, resulting in a responsive and accessible web application that adapts to various devices. Users can navigate the platform efficiently, with features such as floating buttons for WhatsApp communication and real-time feedback to enhance the user experience. Figure 3 illustrates the HealthTest login interface, which includes options to register and access the web application, ensuring seamless interaction for different user roles.

XSL•FO
RenderX

Figure 3. Login interface.



Backend Development

The backend, developed using Spring, was successfully integrated with the frontend, ensuring efficient handling of authentication, data processing, and Application Programming Interface (API) communication. The implementation of RESTful services facilitated seamless data exchange, reducing response times and optimizing server performance. Security features, including authentication and session management, were rigorously tested to ensure data protection.

Database Management

The PostgreSQL database was structured and optimized to efficiently store and manage user data. Query execution times were enhanced through indexing, ensuring rapid retrieval of stored information. Integrity constraints and security measures were implemented to prevent data inconsistencies. Bug reports from external evaluators were used to refine data handling processes, ensuring error-free database transactions and secure user data storage.

The HealthTest web application was successfully deployed as an open-source project, with its source code hosted on GitHub [8, 28], allowing external contributors to review, suggest improvements, and collaborate on new features. Regular updates and bug-fixing processes were implemented to maintain software quality. By adhering to the OSCRUM framework, the project successfully met its development objectives, demonstrating the effectiveness of agile methodologies in open-source software engineering.

Evaluation of OSCRUM's Impact in Digital Health Development

Traditional software development methodologies, such as the waterfall model, often lack flexibility in incorporating real-time user feedback, making them less effective in dynamic fields like mental health. In contrast, OSCRUM allows iterative modifications, ensuring that HealthTest evolves continuously based on user engagement. Unlike proprietary systems that restrict external contributions, OSCRUM promotes an open, transparent development process, making HealthTest an adaptive and scalable tool for emotional health management.

Discussion

Principal Findings

The key findings of this study highlight the development and successful implementation of HealthTest, an open-source web application designed to assess and manage emotional health. The results demonstrate its potential to offer accessible tools for both professionals and general users, promoting mental well-being through reliable psychological resources and tailored recommendations.

The HealthTest web application was developed to support individuals facing mental health challenges. Designed for both professionals and general users, the application provides assistance, advice, and recommendations to improve mental well-being. It is freely accessible without monthly costs, allowing users to assess their emotional state and receive feedback to enhance their current situation. Hosted online, it can be accessed from any internet-connected device. In addition, the application is responsively designed, enabling seamless adaptation to various types of devices.

XSL•FO
RenderX

Several digital health applications have been developed to support emotional well-being. While previous studies, such as MindLogger [12], StudyU [17], and drug-use insights [14], provide valuable contributions, they primarily focus on assessment and monitoring rather than an iterative, open-source development model. HealthTest differs by integrating OSCRUM, which enables continuous community-driven updates and adaptability based on real-world user feedback (See Table 2).

Table . Comparison with existing digital health solutions.

Study	Approach	Limitations	How HealthTest differs
MindLogger [12]	Mobile app for mental health assessments	Limited flexibility, closed-source model	Open-source framework, adaptable to user needs
StudyU [17]	Platform for personalized clinical trials	Requires predefined research protocols	Iterative, real-world usability testing with OSCRUM
DUI [14]	AI-based tool for analyzing substance use language	Focuses only on social media data	Covers broader mental health assessment and intervention

Conclusions

The culmination of this research is reflected in the development of the open-source web application HealthTest, guided by the OSCRUM framework. Comprising 11 stages, this framework provides an agile, organized, and efficient approach to creating open-source web applications. Each step is essential to ensuring software quality and its relevance within the research and development context of HealthTest.

The analysis of primary studies reveals that direct contributions to the development of the application are limited. The field of emotional disorders often focuses on general discussions and systematic studies rather than practical solutions. Despite an exhaustive review, only one study was found to positively integrate into the project [12], while other documents primarily contribute to the theoretical framework.

Considering the insights gathered through the web artifact, a horizon of possibilities emerges for the future. Among the options for expanding HealthTest, a new section focused on diet and exercise is envisioned. This perspective aligns with the scalable and adaptable nature of OSCRUM, enabling the application to address not only emotional aspects but also physical well-being. The incorporation of these sections could have a significant impact on combating challenges associated with emotional disorders.

The development of HealthTest has been a significant step, but it is only the beginning. Continuous exploration and adaptation will be key to maintaining relevance and usefulness in the complex landscape of emotional health.

Acknowledgments

This work was supported by the of the Universidad Técnica Estatal de Quevedo.

Conflicts of Interest

None declared.

References

1. Gómez Conesa A. Impact of the COVID-19 pandemic on mental health symptoms and physiotherapy actions. Fisioterapia 2021(1):1-4. [doi: 10.1016/j.ft.2020.11.001]
2. El Gindi H, Shalaby R, Gusnowski A, et al. The mental health impact of the COVID-19 pandemic among physicians, nurses, and other health care providers in Alberta: cross-sectional survey. JMIR Form Res 2022 Mar 9;6(3):e27469. [doi: 10.2196/27469] [Medline: 34995203]
3. Loza Chiriboga JS, Fiallo López IM, Benítez Serrano AC, Amaya Sigüenza AG. Evaluación de estrés, ansiedad y depresión en ecuador durante la pandemia de COVID-19. Conciencia Social y Salud. 2022. URL: http://revistas.espoch.edu.ec/index.php/cssn/article/view/667/666 [accessed 2024-08-21]
4. Valencia C. Bienestar emocional. La Autoestima November. 2019 Nov 27. URL: https://www.laautoestima.com/bienestar-emocional/ [accessed 2024-08-21]
5. Abrams MP, Pelullo AP, Meisel ZF, Merchant RM, Purtle J, Agarwal AK. State and federal legislators' responses on social media to the mental health and burnout of health care workers throughout the COVID-19 pandemic: natural language processing and sentiment analysis. JMIR Infodemiology 2023;3:e38676. [doi: 10.2196/38676] [Medline: 37013000]
6. Fondo de las Naciones Unidas para la Infancia (UNICEF). La salud mental es determinante para que niños, niñas, adolescentes, familias y comunidades puedan salir adelante. 2020 Oct 20. URL: https://www.unicef.org/ecuador/comunicados-prensa/la-salud-mental-es-determinante-para-que-ni%C3%B1os-ni%C3%B1as-adolescentes-familias-y [accessed 2024-08-21]

XSL•FO
RenderX

7. Carrizo D, Rojas J. Metodologías, técnicas y herramientas en ingeniería de requisitos: un mapeo sistemático. Ingeniare Rev chil ing 2018;26(3):473-485. [doi: 10.4067/S0718-33052018000300473]
8. Rahman SSMM, Mollah SA, Anirban S, et al. OSCRUM: a modified scrum for open source software development. International journal of simulation . [doi: 10.5013/IJSSST.a.19.03.20]
9. Vallejo C. Identificación de aspectos socioemocionales en el contexto del COVID-19.: Ministerio de Educación, Subsecretaría para la Innovación Educativa y el Buen Vivir; 2022. URL: https://tinyurl.com/da2h9867 [accessed 2024-08-21]
10. Ibáñez N. Las emociones en el aula. Estud Pedagog (Valdivia) (28):31-45. [doi: 10.4067/S0718-07052002000100002]
11. Macaluso M, Rothenberg ME, Ferkol T, et al. Impact of the COVID-19 pandemic on people living with rare diseases and their families: results of a national survey. JMIR Public Health Surveill 2024 Feb 14;10:e48430. [doi: 10.2196/48430] [Medline: 38354030]
12. Klein A, Clucas J, Krishnakumar A, et al. Remote digital psychiatry for mobile mental health assessment and therapy: mindlogger platform development study. J Med Internet Res 2021 Nov 11;23(11):e22369. [doi: 10.2196/22369] [Medline: 34762054]
13. Oyri K, Murray PJ. osni.info-Using free/libre/open source software to build a virtual international community for open source nursing informatics. Int J Med Inform 2005 Dec;74(11-12):937-945. [doi: 10.1016/j.ijmedinf.2005.07.023] [Medline: 16112895]
14. Prince Z, Jha D, Singh R. DUI: the drug use insights web server. Bioinformatics 2021 Dec 11;37(24):4895-4897. [doi: 10.1093/bioinformatics/btab461] [Medline: 34164647]
15. Ruan E, Nemeth E, Moffitt R, et al. PLCOjs, a FAIR GWAS web SDK for the NCI prostate, lung, colorectal and ovarian cancer genetic Atlas project. Bioinformatics 2022 Sep 15;38(18):4434-4436. [doi: 10.1093/bioinformatics/btac531] [Medline: 35900159]
16. Dobbins NJ, Spital CH, Black RA, et al. Leaf: an open-source, model-agnostic, data-driven web application for cohort discovery and translational biomedical research. J Am Med Inform Assoc 2020 Jan 1;27(1):109-118. [doi: 10.1093/jamia/ocz165] [Medline: 31592524]
17. Konigorski S, Wernicke S, Slosarek T, et al. StudyU: a platform for designing and conducting innovative digital N-of-1 trials. J Med Internet Res 2022 Jul 5;24(7):e35884. [doi: 10.2196/35884] [Medline: 35787512]
18. Motahar SM, Safie N, Mukhtar M, Aljunid SM, Mostafavi S. An applied approach to teach hospital information systems development using an open source ERP framework. Procedia Technology 2013;11:1259-1265. [doi: 10.1016/j.protcy.2013.12.322]
19. Vidal-Balea A, Blanco-Novoa Ó, Fraga-Lamas P, Fernández-Caramés TM. Developing the next generation of augmented reality games for pediatric healthcare: an open-source collaborative framework based on ARCore for implementing teaching, training and monitoring applications. Sensors (Basel) 2021 Mar 7;21(5):1-24. [doi: 10.3390/s21051865] [Medline: 33800070]
20. de la Iglesia I, Vivó M, Chocrón P, Maeztu GD, Gojenola K, Atutxa A. An open source corpus and automatic tool for section identification in Spanish health records. J Biomed Inform 2023 Sep;145:104461. [doi: 10.1016/j.jbi.2023.104461] [Medline: 37536643]
21. Silva D. Bug reporting template. Google Docs. 2024. URL: https://docs.google.com/document/d/14uSPVDNt3gS2ydSFiJNlTQgyCw60CxbI/edit?tab=t.0 [accessed 2024-08-08]
22. Boada Oriols M, Gómez Gutiérrez JA. El Gran Libro de Angular: MARCOMBO, S.A; 2018. URL: https://www.academia.edu/81935666/El_gran_libro_de_Angular [accessed 2024-08-21]
23. Christudas B. MySQL. In: Practical Microservices Architectural Patterns: Apress, Vol. 2019:877-884. [doi: 10.1007/978-1-4842-4501-9_27]
24. Silva D. A web artifact for feedback collection. Blogger. 2024. URL: https://bienestaremocionalysaluds.blogspot.com/2022/11/ansiedad-estres-tristeza.html [accessed 2024-08-08]
25. Silva D. Results of information-gathering techniques for healthtest. Google Drive. 2024. URL: https://drive.google.com/file/d/1z-dPTo5vLAJsd18eIfxgJiVKPhxstp7C/view [accessed 2024-08-08]
26. Silva D. General use case diagram for healthtest. Google Drive. 2024. URL: https://drive.google.com/file/d/1sIfQFDKyhm5DPQOIE673-rDYaI8UPq9j/view [accessed 2024-08-08]
27. Silva D. Complete bug reports for healthtest. Google Drive. 2024. URL: https://drive.google.com/file/d/1W1BoSyHmMNxwWaZX-g7BtzL1Cl8T0Whv/view [accessed 2024-08-14]
28. Silva D. HealthTest: source code repository. GitHub. 2024. URL: https://github.com/dsilvas2001/HealthTest [accessed 2024-08-14]

Abbreviations

ANGULAR/CLI: Angular Command Line Interface
ERP: Enterprise Resource Planning
OSCRUM: Open Source Scrum
SMS: Systematic Mapping Study

XSL•FO
RenderX

Edited by E Mensah; submitted 28.11.24; peer-reviewed by JW Castro, S Siddiqui; revised version received 05.02.25; accepted 14.06.25; published 11.08.25.

Please cite as:
Llerena L, Nata Castro D, Rodriguez N, Silva Sánchez D
Open Source Web Application (HealthTest) for Emotional Health and Wellness Management in University Students: Development and Usability Study
Online J Public Health Inform 2025;17:e69413
URL: https://ojphi.jmir.org/2025/1/e69413
doi:10.2196/69413

© Lucrecia Llerena, Daisy Nata Castro, Nancy Rodriguez, Donald Silva Sánchez. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 11.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying Substance Use and High-Risk Sexual Behavior Among Sexual and Gender Minority Youth by Using Mobile Phone Data: Development and Validation Study

Mehrab Beikzadeh[1], MS; Ian W Holloway[2], MPH, MSW, PhD; Kimmo Kärkkäinen[3], PhD; Chenglin Hong[4], PhD; Cory Cascalheira[5], PhD; Elizabeth S C Wu[2], MPH; Callisto Boka[6], BA; Alexandra C Avendaño[2], MA; Elizabeth A Yonko[6], MPH; Majid Sarrafzadeh[1], PhD

[1]Department of Computer Science, UCLA Samueli School Of Engineering, University of California, Los Angeles, 7400 Boelter Hall, Los Angeles, CA, United States

[2]Department of Social Welfare, University of California, Los Angeles, Los Angeles, CA, United States

[3]Optum, Los Angeles, CA, United States

[4]School of Social Work, University of Connecticut, Hartford, CT, United States

[5]Addiction Treatment Center, VA Puget Sound Health Care System, Seattle, WA, United States

[6]Department of Epidemiology, UCLA Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Mehrab Beikzadeh, MS
Department of Computer Science, UCLA Samueli School Of Engineering, University of California, Los Angeles, 7400 Boelter Hall, Los Angeles, CA, United States

Abstract

Background: Sexual and gender minority (SGM) individuals are at heightened risk for substance use and sexually transmitted infections than their non-SGM peers. Collecting mobile phone usage data passively may open new opportunities for personalizing interventions, as behavioral risks could be identified without user input.

Objective: This study aimed to determine (1) whether passively sensed mobile phone data can be used to identify substance use and sexual risk behaviors for sexually transmitted infection (STI) and HIV transmission among young SGM who have sex with men, (2) which outcomes can be predicted with a high level of accuracy, and (3) which passive data sources are most predictive of these outcomes.

Methods: We developed a mobile phone app to collect participants' messaging, location, and app use data and trained a machine learning model to predict risk behaviors for STI and HIV transmission. We used Scikit-learn to train logistic regression and gradient boosting classification models with simple linear model specification to predict participants' substance use and sexual behaviors (ie, condomless anal sex, number of sexual partners, and methamphetamine use), which were validated using self-report questionnaires. F_1-scores were used to quantify prediction accuracy of the model using different data sources (and combinations of these sources) for prediction. Differences between text, location, app use, and Linguistic Inquiry and Word Count (LIWC) domains by outcome were investigated using independent *t* tests where associations were considered significant at $P<.05$.

Results: Among participants (n=82) who identified as SGM, were sexually active, and reported recent substance use, our model was highly predictive of methamphetamine use and having ≥6 sexual partners (F_1-scores as high as 0.83 and 0.69, respectively). The model was less predictive of condomless anal sex (highest F_1-score 0.38). Overall, text-based features were found to be most predictive, but app use and location data improved predictive accuracy, particularly for detecting ≥6 sexual partners. Methamphetamine use was significantly associated with dating app use ($P=.01$) and use of sex-related words ($P=.002$). Having ≥6 sex partners was associated with dating app use (0.02), use of sex-related words ($P=.001$), and traveling a further distance from home ($P=.03$), on average, compared to participants with fewer sex partners. Methamphetamine users were more likely to use social ($P=.002$) and affect words ($P=.003$) and less likely to use drive-related words ($P=.02$). People having 6 or more partners were more likely to use social, affect words, and cognitive process-related words ($P=.003$ and .004 respectively).

Conclusions: Our results show that passively collected mobile phone data may be useful in detecting sexual risk behaviors. Expanding data collection may improve the results further, as certain behaviors, such as injection drug use, were quite rare in the study sample. These models may be used to personalize STI and HIV prevention as well as substance use harm reduction interventions.

International Registered Report Identifier (IRRID): RR2-10.2196/58448

XSL•FO
RenderX

(Online J Public Health Inform 2025;17:e68013) doi:10.2196/68013

KEYWORDS

substance use; HIV risk; sexual and gender minoritized; mobile app; eHealth

Introduction

Sexual and gender minoritized (SGM) individuals are at heightened risk for substance use and sexually transmitted infections (STIs) than the general United States population. Among SGM populations, men who have sex with men (MSM), for example, are twice as likely to use illicit drugs [1], which may be used to cope with negative life events and thoughts, or to enhance pleasure during sex [2]. Over half of new HIV infections occur among SGM, which can be attributed to sexual risk behaviors and intravenous drug use (IDU) [3-5]. Between 2018 and 2022, the Centers for Disease Control and Prevention reported HIV diagnoses increased significantly among transgender and gender nonbinary populations, more so than among cisgender men or women [5]. Research suggests that these health disparities in substance use and HIV are generated by unjust social conditions [6,7] and increased exposure to minority stressors [8,9]. SGM are also at higher odds of mental distress and depression [10,11], which in turn may increase substance use as a coping mechanism [2,12].

Systematic reviews of studies in SGM populations, largely thus far tailored for MSM, have shown that interventions can be effective on methamphetamine- and sexual health-related outcomes, such as condomless anal sex or substance use during sex [13], and participants find these interventions useful for gaining new knowledge and skills [14]. In addition, participants find interventions useful for self-reflection [14], which may lead to behavior change. However, results from a global survey among MSM who use substances found that only 11% of respondents had access to substance use treatment programs and only 5% participated in such a program [15]. In the United States, only 6.5% of people who needed substance use treatment received it in 2020 [16]. The majority of those who want substance use treatment but do not receive it experience significant access barriers such as affordability due to the lack of health care coverage, not finding an appropriate program, fear of others having a negative opinion of them, and the absence of culturally informed treatment tailored to the unique needs of SGM [17]. Therefore, efforts to address these disparities should prioritize improving access to health services for SGM, particularly strategies to deliver health services and identify those at the highest risk.

Mobile- and eHealth-based interventions could improve the accessibility of interventions, as they have potential to overcome many of these treatment barriers (eg, overcoming the stigma of receiving substance use treatment by using the eHealth intervention from the privacy of one's home). Mobile and eHealth interventions may also open new opportunities for personalization through increased availability of data about participants. Prior studies have shown success in providing personalized HIV interventions to MSM and people using substances [18-20]. However, this personalization typically depends on participants reporting behaviors manually, which increases participant burden. For example, one study asked participants to respond to either daily or biweekly surveys, which many participants reported to be too repetitive [21]. Although burdensome, the group receiving daily surveys found them to be more useful than the group receiving biweekly surveys. This indicates that behavioral health monitoring should not depend on receiving frequent input from the participant. Therefore, being able to automate some or all the behavior monitoring could reduce participant burden.

In this study, we investigate how machine learning techniques can help identify sexual risk behaviors among SGM from passively sensed mobile phone data. Prior studies have predicted HIV risk using, for example, Twitter [22], electronic health records [23,24], or smartphone survey data [25]. Similarly, substance use risk has been detected using survey data [26], cognitive test results [27], Instagram (Meta) profile data [28], and social media posts [29]. To the best of our knowledge, this is the first study to correlate substance use and sexual risk behaviors among SGM using passively collected mobile phone data, which allows for frequent data collection with minimal effort required from the participant.

We first developed a mobile sensing app that tracks participants' daily actions, such as their location, messaging, and app use. We then trained machine learning models to detect substance use and sexual risk behaviors from these data and evaluated their performance in predicting different behaviors. Finally, we analyzed how different risky behaviors manifest in mobile phone data.

The main contributions of this study are (1) demonstrating how passively collected mobile phone data can be used for behavioral risk prediction and identifying limitations of this approach; (2) evaluating which types of data should be collected to identify substance use and sexual risk behaviors by training machine learning models using different subsets of the data, as well as analyzing differences between participants' data; and (3) Determining how accurately different behaviors can be identified from mobile phone data.

Methods

Design and Eligibility

Data for this analysis were derived from a National Institutes of Health–funded randomized comparison trial – uTECH (ClinicalTrials.gov identifier: NCT04710901). To be eligible for the study, participants had to meet the following criteria: (1) be 18 to 29 years old, (2) be able to speak in English, (3) identify as a sexual or gender minority, (4) have had anal or oral sex with a man in the past 3 months, (5) have used substances (such as alcohol, marijuana, poppers [amyl nitrate], methamphetamines, heroin, cocaine, and ecstasy) in the past 3 months, (6) have had sex while using substances in the past 3 months, (7) being HIV negative or of unknown HIV status, (8) have used a dating app to meet sexual and substance use partners

XSL•FO
RenderX

in the past 3 months, (9) own a smartphone, (10) reside in the United States, (11) be willing to participate in a 12-month study, and (12) be able to provide informed consent. Eligibility criteria, recruitment procedures, and overall study design are detailed comprehensively in the protocol paper [30].

Screening

All participants completed an initial screener survey that was hosted on Qualtrics, a web-based survey platform [31]. The screener provided information about the study and included questions to determine eligibility. If an individual met eligibility criteria, the survey used branching logic to show additional screens to ask for contact information, including the phone number of their smartphone [30].

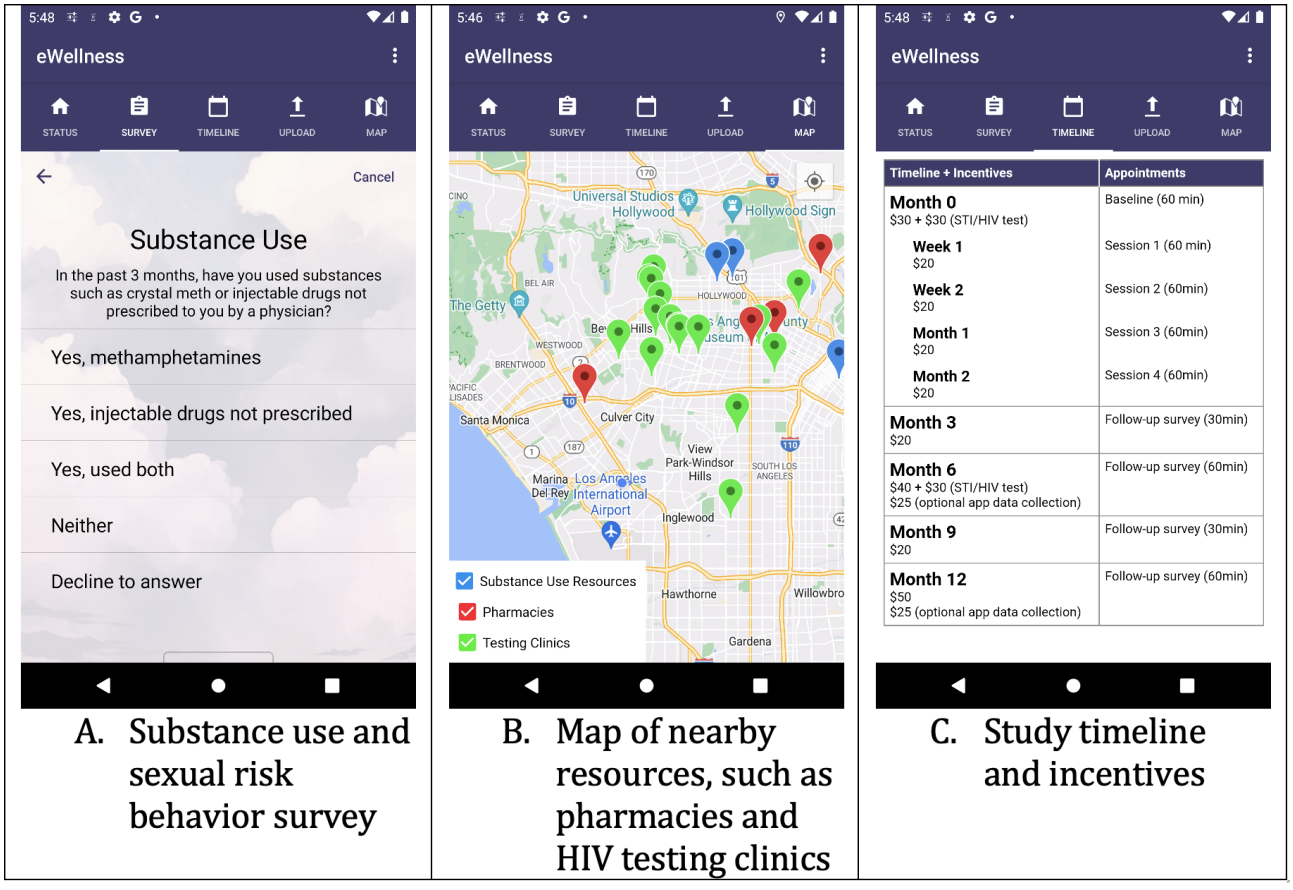
Research staff took precautions against fraudulent screeners by using survey metadata to identify noncellular phone numbers, virtual private network software, and high-risk IP addresses [30]. Screeners that were suspected to be illegitimate were removed before enrollment in the study. SGM who completed the screener, met the eligibility criteria, and passed fraud detection checks were contacted to schedule a consent and onboarding session over the Zoom conferencing platform (Zoom Video Communications) [32].

Ethical Considerations

All study procedures and protocols were approved by the by the South Campus institutional review board of the University of California, Los Angeles (IRB#22 - 000009). Informed consent was obtained during the onboarding process. During the informed consent process, the interviewer shared the consent document which provided details of what types of data (eg, keylogged data and Global Positioning System [GPS] data) were collected by the data collection app. The consent document also provided details on what kinds of data would not be collected (ie, photos and video). Benefits and risks of participation, monetary incentives, and data privacy protections were likewise detailed in the consent document. Participants were informed all data collected were protected from use as evidence in legal processes by a Certificate of Confidentiality granted by the National Institutes of Health (number: CC-OD-22 - 3555). If the participant consented to participate, their agreement was recorded by the interviewer, and they were emailed a copy of the consent to keep for their records. Participants were compensated up to US $450 for completion of study-related activities (see Figure 1C for details).

All research staff were required to complete HIPAA (Health Insurance Portability and Accountability Act) and human subjects research training to gain access to participant materials. Any personally identifiable information or media collected unintentionally through data collection was redacted or removed entirely prior to storage in the study's database.

Figure 1. Screenshots of the eWellness app.



A. Substance use and sexual risk behavior survey

B. Map of nearby resources, such as pharmacies and HIV testing clinics

C. Study timeline and incentives

XSL•FO
RenderX

Data Collection App

Android [33] users who completed the consent and enrollment process would then install the data collection app used in the study, eWellness. The app was based on the Aware Framework [34], which has been used in numerous earlier eHealth studies, for example, to predict depression and anxiety [35,36], progression of Parkinson disease [37], or alcohol use events [38]. We adapted eWellness for Android phones to collect data on participants' mobile phone use activities. We asked participants to give the app all the necessary permissions to passively collect keyboard and location data during participation. In addition to collecting keyboard data when participants typed text, the app collected information on which app they were typing the text in or, if they were using a browser, which website they were on.

The app also contained a substance use and sexual behavior survey, referred to as the "wellness survey," (shown in Figure 1A) which the participants were asked to complete when they joined the study and once every 3 months after that. The survey was adapted from the US Centers for Disease Control and Prevention's HIV and pre-exposure prophylaxis (PrEP) clinical practice guidelines [39]. The wellness survey contained questions on individual participants' substance use and sexual risk behaviors, which yielded a total score indicating one's risk for HIV infection and PrEP eligibility. The questions and answer options are shown in Table S1 in the Multimedia Appendix 1. To keep participants engaged, the app also provided other useful information, such as a map of nearby resources, such as pharmacies, HIV testing locations, and substance use harm reduction resources (Figure 1B), as well as the study timeline and incentives (Figure 1C).

The app sent the collected data to our secure server every 30 minutes whenever internet connection was available. Highly sensitive information, such as passwords, was filtered out and the research team had no access to them. The server-stored data did not contain other identifying information; only a randomly assigned participant identifier was included in the data. A document linking personal information necessary for participant follow-up was hosted on Federal Information Processing Standards 140‐2 certified cloud-based file management platform, Box. The Box directory was only accessible via multifactor authentication university-based single sign-on log-in to institutional review board–approved researchers with participant follow-up duties who had completed HIPAA and Good Clinical Practices trainings. All files containing participant information were protected with AES 256-bit encryption and data leak prevention and threat detection algorithms integrated into Box.

Data Preprocessing

Keyboard Data

The eWellness app collected information about the currently active text field's contents on every keystroke. As a result, our database contained multiple rows of data for every full line of text that the participant typed. For example, typing "Hello" might have been stored in the database as rows containing text values: "H," "He," "Hel," "Hell," and "Hello." In addition, the participant could have changed earlier parts of the text or used autocorrection, which means that this same text could have appeared as database rows: "H," "He," "Hel," "Helo," "Hello." As a result, the earlier row was not always a substring of the next one.

To remove duplicate rows, we repeated the following steps for each individual participant's text data until there were no more rows to remove: (1) compare each row to the next row and if the first row is a substring of the second one, remove the first row, and (2) calculate the Levenshtein similarity between each row and the following row, and if the similarity is larger than 0.6, remove the first row.

Levenshtein similarity between two strings a and b is defined as:

$$(1)\quad sim(a,b)=1-\frac{dist(a,b)}{max(len(a),len(b))}$$

where dist(a,b) is the Levenshtein distance [40] which counts the minimum number of single-character modifications (insert, delete, substitute) that are necessary to make the strings identical:

$$(2)\quad dist(a,b)=\begin{cases}len(a), & \text{if } len(a)=0\\ len(b), & \text{if } len(b)=0\\ dist(tail(a),tail(b)), & \text{if } a[0]=b[0]\\ 1+min\begin{cases}dist(tail(a),b)\\ dist(a,tail(b))\\ dist(tail(a),tail(b))\end{cases} & \text{otherwise}\end{cases}$$

where *a[0]* refers to the first character of string *a*, and *tail(a)* refers to a substring of *a* which contains everything except the first character. We calculated the similarity score using the TextDistance Python library [41].

Application Data

For every row of text data, we had a package name of the app where the text was entered as well as the Uniform Resource Locator (URL) of the website if the participant was using a web browser. In many cases, online services can be accessed both through an app and through a website, so we combined these data sources by extracting the domain name from the URL and mapping the commonly appearing URLs to the corresponding package names using a manually curated list of domain name or app pairs. If a domain name was not in this list, the domain name itself was used as the package name. Apps and websites were treated in the same manner in analysis and model training.

Location Data

The eWellness app saved GPS coordinates periodically whenever the phone moved to a different location. The app avoided unnecessary data collection to reduce battery consumption by only collecting location data when the phone was moving. This meant that if the participant remained in the same location, we did not receive location data until the participant started moving again. As we were only interested in locations in which the participant spent time rather than locations that the participant moved by, we removed data points where the participant was moving and only retained the last location once the movement ended.

Feature Extraction

After cleaning the dataset using the previously shown steps, we manually extracted various features (ie, computed independent variables) from each of the data sources to be used with the machine learning algorithms. These feature extraction techniques are described in the following subsections.

XSL•FO
RenderX

Text Data

In our dataset, participants were active for different numbers of days, and for individual participants, different days had sometimes vastly different amounts of text data. This made the direct application of traditional text processing techniques challenging. In addition, the words and phrases used by participants sometimes differed from the ones used by the general public, so for optimal results, the techniques had to be tailored for the study population. We only considered text data collected from social media, dating, or messaging apps/websites, as text data from other sources was found to contain more noise than useful information (for example, product names in shopping apps or location names in navigation apps).

The first set of features extracted from the text data was the frequencies of individual words used. Participants' text was first lemmatized, which means that inflected word forms were transformed to their base forms (eg, "walking" → "walk," "better" → "good") to avoid having the same word appear in multiple forms in our set of features. Lemmatization was performed using WordNet Lemmatizer from the Natural Language Toolkit (NLTK) [42]. Then, we removed words defined in NLTK's stop word list, commonly appearing placeholder texts (eg, "Enter message," "Say something"), as well as words used by fewer than five people. For each remaining word, we calculated the frequency of word use:

(3)freq(w)=# days with word w# days with text data

The second set of text features only considered words and phrases associated with drug use or sexual behaviors. We used a phrase list, which had been found effective for identifying HIV risk behavior as well as substance use by an earlier study [29], and we used our previously defined frequency formula to determine frequencies for both individual phrases as well as for higher-level phrase categories (ie, different types of substance use or sexual behavior).

A third set of features was computed by Linguistic Inquiry and Word Count (LIWC) software [43], which uses built-in dictionaries to capture social and psychological states. It computes features describing how much an individual talks about a variety of topics, such as money, physical intimacy, or leisure activities, and it also computes higher-level descriptive features to measure factors, such as analytical thinking, authenticity, and emotional tone. It has been used in numerous studies to, for example, analyze fake news [44], social media posts [45,46], online reviews [47], and college admission essays [48]. We used it to generate features for each individual day, and we calculated the average across all days for each individual participant.

Our last set of text features was generated using the Bidirectional Encoder Representations from Transformers (BERT) language model [49]. We used a model that was pretrained for sentiment analysis using Twitter data [50], as we expected Twitter data to use similar language as other social media and messaging platforms. We removed the last fully connected layer of the model so that it could be used to generate text embeddings, and we applied it to each individual day of data. These text embeddings were then averaged across all days of data for individual participants.

App Data

We captured app usage by looking at apps where the user wrote text. We generated one set of features by calculating how frequently each app was used:

(4)freq(app)=# days using app# days using any apps

We also considered a subset of these frequency features that only contained social media, dating, and messaging apps, as we expected other types of apps to be less relevant to our prediction task. Other apps (eg, maps, music, or shopping) were expected to be noisy and therefore to have a negative effect on the model's predictive performance.

Location Data

Before extracting features from location data, we clustered GPS coordinates for each individual participant by using the Mean-Shift algorithm [51]. This algorithm moves all points repeatedly towards the mean value of their neighborhood (determined by window radius r) until all points have converged. Points that converge to the same coordinates are defined to belong in the same cluster, thus allowing the algorithm to find the appropriate number of clusters. As the generated clusters depend on the window size, we determined the appropriate size by visually inspecting the clustering results. We also assumed that the most visited location was the participant's home.

We then computed the features describing individuals' mobility, such as how far from home they traveled, how many locations they visited per day, and how many of these locations were unique. These features were selected such that they were potentially related to participants' behavioral health outcomes either directly or indirectly. For example, several unique locations may be associated with having many partners, while having very few unique locations could be related to methamphetamine use due to the limited number of locations where the participant could safely use the drug. The full list of location-based features is shown in Multimedia Appendix 1.

Model Training

We used Scikit-learn [52] to train logistic regression [53] and gradient boosting [54] classification models to predict participants' answers to each survey question. These models were chosen to represent a simple linear model as well as a more advanced nonlinear model. To determine which types of data could be useful for the prediction task, we trained separate models using individual data categories, such as location data, app use, and risky word use. We then evaluated combinations of these features, focusing on feature combinations that we believed would give a comprehensive view of the participant's activities without including redundant data (eg, not including social media apps and all apps in the same model). Models were evaluated using leave-one-out cross-validation due to the relatively small number of participants.

To address class imbalance, we calculated F_1-scores for the minority class and used gradient boosting which can better handle imbalanced datasets. After initial exploration of model hyperparameters, we selected values that provided stable

XSL•FO
RenderX

performance. For logistic regression, we used default hyperparameters with max_iter=1000 to ensure convergence. For gradient boosting, we used a GradientBoostingClassifier with n_estimators=80 and default values for other hyperparameters.

We chose to focus on F_1-scores for model evaluation, as they balance precision and recall considerations. In the context of behavioral risk prediction for potential interventions, both types of misclassification errors have important implications: false positives might lead to unnecessary interventions, while false negatives could miss individuals who might benefit from support. The F_1-score helps balance these considerations for our exploratory analysis, though future applications may need to adjust classification thresholds based on specific intervention contexts.

Results

Study Population

Sociodemographic, sexual risk, and substance use characteristics reported by participants (n=82) at baseline are summarized in Table 1.

XSL•FO
RenderX

Table . Self-reported participant characteristics at baseline.

Characteristic	Total
Participants, n (%)	82 (100)
Age, mean (SD)	25.2 (3.9)
Race and ethnicity, n (%)	
American Indian or Native Alaskan	2 (2.4)
Asian	13 (15.9)
Black or African American	7 (8.5)
Hispanic or Latino	10 (12.2)
Middle Eastern and North African	1 (1.2)
Two or more races	11 (13.4)
Non-Hispanic White	38 (46.3)
Gender identity, n (%)	
Cisgender man	54 (65.9)
Transgender man	14 (17.1)
Nonbinary	11 (13.4)
Transgender woman	3 (3.7)
Sexual orientation, n (%)	
Gay	55 (67.1)
Bisexual or Pansexual	20 (24.4)
Queer	5 (6.1)
Straight or heterosexual	1 (1.2)
Refuse to answer	1 (1.2)
Education, n (%)	
Less than college degree	33 (40.2)
College degree or higher	48 (58.5)
Refuse to answer	1 (1.2)
US region, n (%)	
West	26 (31.7)
Northeast	25 (30.5)
South	17 (20.7)
Midwest	14 (17.1)
Sexual behavior, n (%)	
Condomless receptive sex	61 (74.4)
Condomless insertive sex with HIV+ partner	3/74 (4.1)
5+ times	
HIV+ partners	8/74 (10.8)
6+ partners	41 (50)
11+ partners	26 (32)
Substance use, n (%)	
Methamphetamine use	15 (18.3)
Injection drug use	3 (3.7)
Injects cocaine	1 (1.2)
Injects in group	2 (2.4)

XSL•FO
RenderX

Characteristic	Total
Shares injection equipment	1 (1.2)
Injects methamphetamine	2 (2.4)
In substance use treatment program	0 (0)

Data Statistics

We collected data from participants between November 10, 2021, and April 15, 2024. Dataset statistics are shown in Table 2. Among all the apps, we manually identified 68 social media, dating, and messaging apps which were later used to analyze participants' messaging data. It should be noted that apps included unique websites as well (grouped by domain name).

We used data for all participants who had at least 30 days of data available. If a participant's answer to a survey question was "Decline to answer" or "I don't know," this answer was not included in the model training or evaluation. This did not, however, exclude their other survey answers from being used. Statistics for survey responses are shown in Table 1 (full questions and answer options are shown in Table S2 in Multimedia Appendix 1).

Table . Text, location, and app use summary statistics.

	Total	Mean	Median	SD	Min	Max
Lines of text	4,848,639	59,129.7	45,758	43,832.2	2529	195,797
Locations	1,607,149	16,917.4	6071	26,793	509	169,334
Unique apps	2248	92.1	89.5	46.4	20	289

Model Performance

We trained classification models to predict answers to each question. As some questions had partially overlapping answer options, we split them into multiple distinct questions. For example, the answer options for substance use included methamphetamine use, injection drug use, and both, so we split it into two questions: methamphetamine use and injection drug use. In addition, some questions had a very low number of positive responses (for example, only 2 participants were in a substance use treatment program). As a result, the focus of our discussion will be on the three questions which we determined to be the most informative: methamphetamine use, having 6 or more male sexual partners, and receptive anal sex without a condom.

Results for predicting survey responses using both individual feature types as well as combinations of them are shown in Table 3. Feature combinations were selected both based on their individual results and based on whether they were presumed to provide nonoverlapping information. For example, we avoided combining highly correlated feature groups, such as social media apps and all apps, in the same model.

As the results show, methamphetamine use could be predicted well using just the text data. The word frequency feature with the gradient boosting model worked best. Predicting having many partners worked reasonably well when combining all feature types. Predictive models were only moderately successful in determining whether the participant had receptive condomless anal sex.

Combining multiple feature types rarely improved the performance by a noticeable amount. This could be because in many cases, the feature groups might provide redundant information, so using only one highly informative feature group was enough. In addition, increasing the number of features could lead to overfitting, as the number of features can become much larger than the number of participants.

XSL•FO
RenderX

Table . F_1-scores for predicting answers to survey questions. F_1-score was calculated for the less frequent response, which in most cases was the "positive" answer (answer frequencies are shown in Table 1). The first value shows the score using logistic regression and the second value shows the score using a gradient-boosting classifier.

	Methamphetamine use	Sexual behavior	
		6+ partners	Condomless receptive sex
Social apps	0.32/0.40	0.49/0.53	0.35/0.16
All apps	0.31/0.08	0.56/0.60	0.29/0.33
Location	0.00/0.25	0.46/0.39	0.00/0.00
Risky words	0.48/0.52	0.62/0.65	0.29/0.12
All words	0.29/0.83[a]	0.57/0.23	0.22/0.11
LIWC[b]	0.52/0.47	0.61/0.63	0.31/0.26
Bert	0.34/0.34	0.67/0.64	0.30/0.38[a]
Risky words, LIWC	0.50/0.67	0.62/0.61	0.32/0.25
Social apps, Risky words	0.48/0.38	0.58/0.65	0.28/0.15
Social apps, Risky words, LIWC	0.48/0.67	0.60/0.55	0.33/0.31
Social apps, Risky words, Location	0.52/0.48	0.61/0.69[a]	0.24/0.17
Social apps, Risky words, Location, LIWC,	0.52/0.67	0.61/0.64	0.19/0.07
All	0.42/0.58	0.62/0.49	0.10/0.21

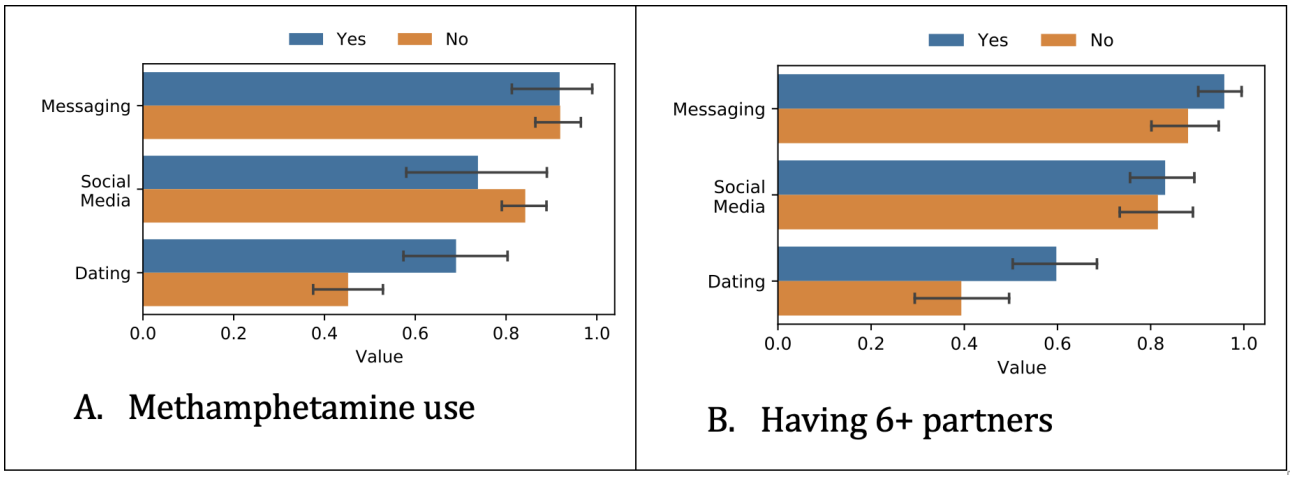
[a]Highest predictive result for the corresponding outcome.

[b]LIWC: Linguistic Inquiry and Word Count

Feature Analysis

Next, we analyzed how the participant data differed depending on the survey responses. We show the differences based on independent *t* tests for the most predictive tasks (Figures 2-4), which were methamphetamine use and having ≥6 partners.

Figure 2. Differences in app use among different groups. The x-axis represents the percentage of days when the participant communicated using an app from a certain category. The blue bar corresponds to "Yes" methamphetamine use or "Yes" had ≥6 partners, respectively.



XSL•FO
RenderX

Figure 3. Differences in risky word use among different groups. The x-axis represents the percentage of days when the participant used words or phrases from a certain category. The blue bar corresponds to "Yes" methamphetamine use or "Yes" had ≥6 partners, respectively.

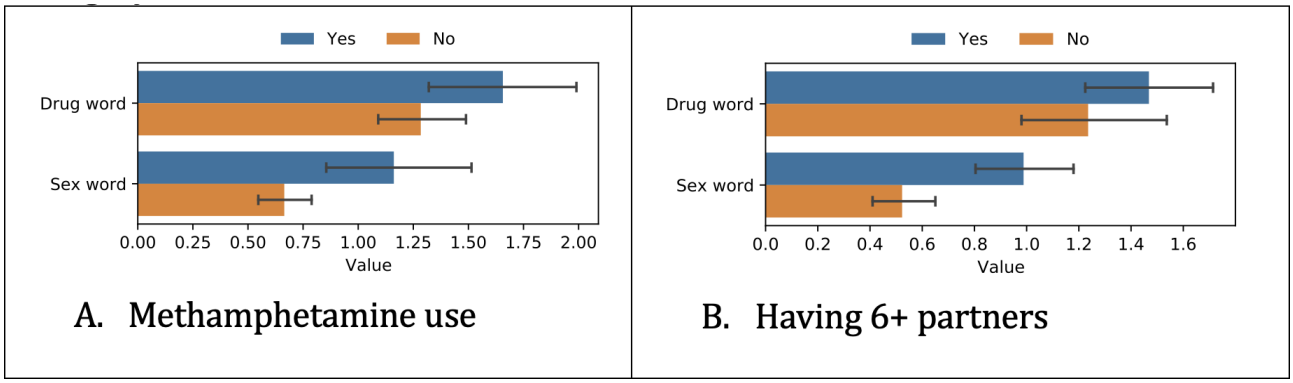
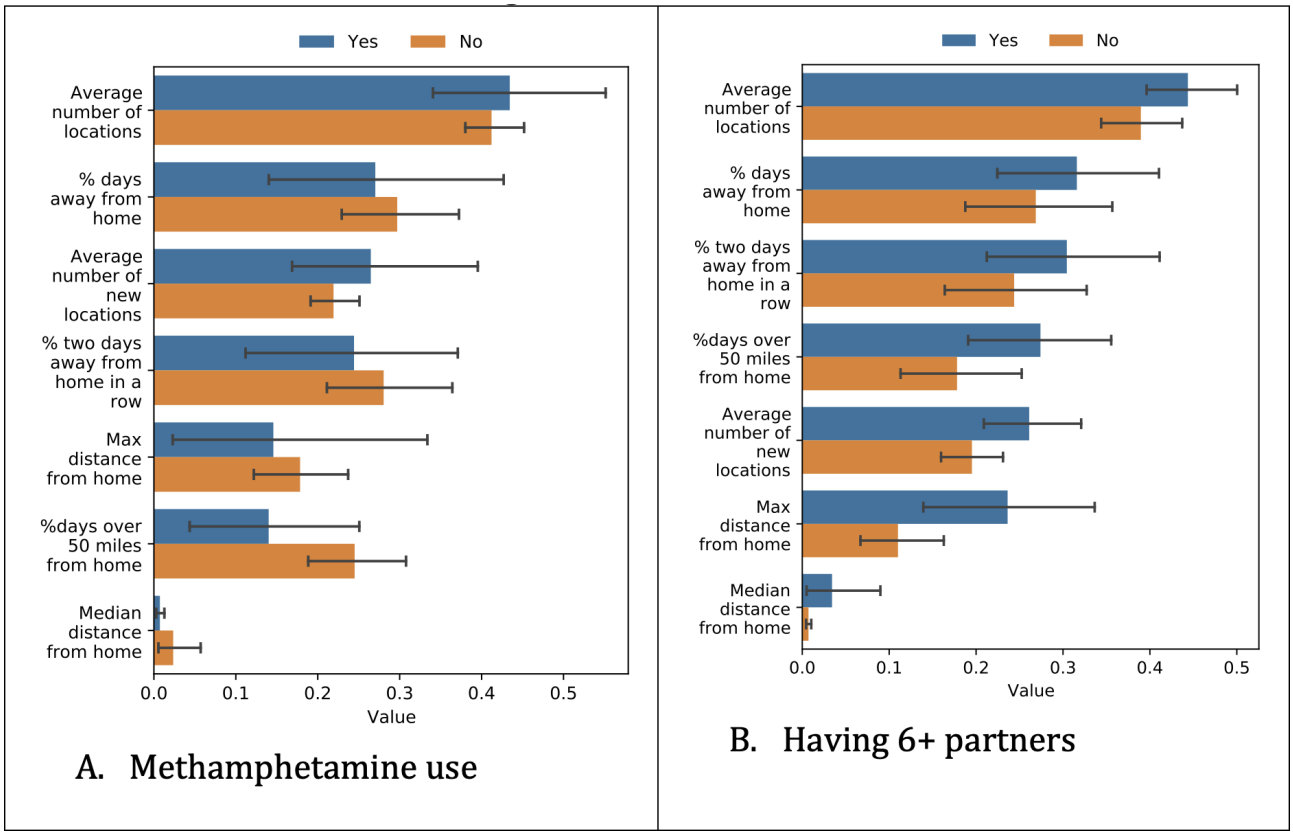


Figure 4. Differences in location data among different groups. Values have been scaled such that the largest individual value for each feature becomes 1 to be able to show all values in the same figure. The blue bar corresponds to "Yes" methamphetamine use or "Yes" had ≥6 partners, respectively.



App Use

Figure 2 shows how frequently participants used apps from different categories. We considered any apps that are used for communicating with other people and divided them into 3 categories: messaging apps (eg, Messages, WhatsApp, and Telegram), social media apps (eg, Facebook, Instagram, Twitter [currently known as X]), and dating apps (eg, Grindr, Tinder, and Adam4Adam).

Methamphetamine users or participants who had ≥6 partners were more likely to use dating apps than nonusers (P=.01, and P=.02, respectively). However, differences in the use of social media and messaging apps were not statistically significant for these groups.

Risky Words

Figure 3 shows differences in risky word use. We divided the list of risky words into sex-related and drug-related words.

Methamphetamine users and individuals with 6+ partners were both more likely to use sex-related words, with significance levels of P=.002 and P=.001, respectively. However, the difference in drug-related word usage was not statistically significant for methamphetamine users and individuals with 6+ partners. (P=.12 and .21, respectively)

Location

Figure 4 shows how location data differed for different groups. Due to the large number of location-based features, we chose a smaller subset of features that contained less overlapping information.

XSL•FO
RenderX

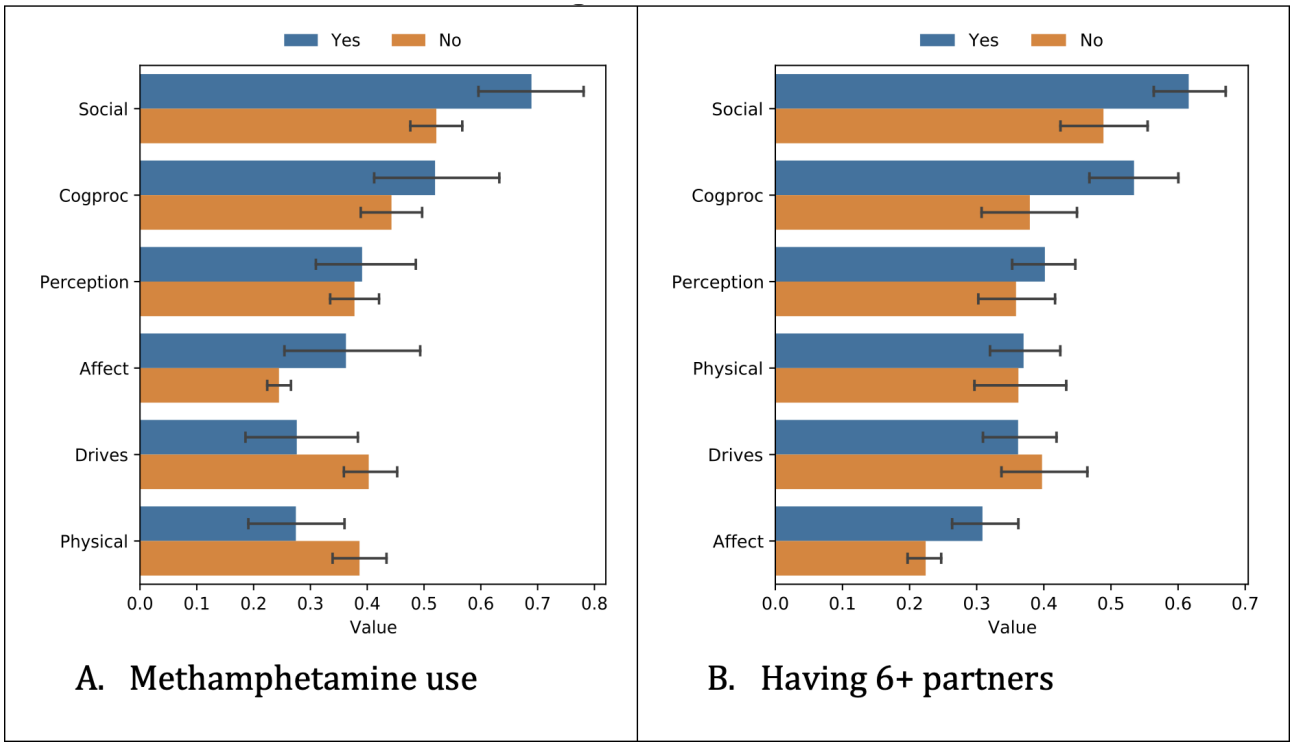
Methamphetamine users were less likely to spend time over 50 miles from home, although the difference was not statistically significant (*P*=.14). People with 6+ partners were found to travel further than those with 5 or fewer partners (*P*=.03).

LIWC

Lastly, we compared LIWC features among different groups (Figure 5). Again, due to the large number of distinct features, we show results only for some of the super-categories which we expected to show differences.

Methamphetamine users were more likely to use social (*P*=.002) and affect words (*P*=.003) and less likely to use drive-related words (*P*=.02). People having 6 or more partners were more likely to use social, affect words, and cognitive process-related words (*P*=.003 and .004 respectively).

Figure 5. Differences in Linguistic Inquiry and Word Count features among different groups. Original values have been scaled to fit in the same figure.



Discussion

Principal Results

In this paper, we have shown that mobile sensing data can be used to identify multiple risk behaviors of SGM in our study sample. More specifically, our participants' text and location data were highly informative of methamphetamine use and having over 6 sexual partners in three months.

In addition to determining which behaviors can be predicted, our second goal was to determine what data is useful for these predictions. We have shown that text-based features like all words, risky words, and BERT were the most informative for most behaviors, which was an expected result because participants might, for example, look for partners on dating apps or discuss substance use in private messages with other people. This is aligned with previous research [29] showing that certain types of substance use may be predicted from social media messaging data.

In addition, we have shown that more recent language modeling techniques, such as BERT, can often provide similar results as the traditional techniques based on predetermined word lists and word frequencies. However, the more abstract nature of these representations may complicate the interpretation of the results, as individual values do not have a human-interpretable meaning. This lack of human interpretability might not be a limitation in digital health applications, where the emphasis is on achieving high precision and recall for effective intervention delivery rather than on understanding the underlying phenomena. On the other hand, BERT representations can also help improve privacy, as they do not reveal which exact words the participants used. Due to the small number of participants, training a language model using our dataset was not feasible, so we relied on a model that had been trained on Twitter (currently known as X) data. While we expect the language use to be mostly similar across datasets, training a language model using data from the target population could improve the results if enough data were available, as people might use different words and phrases on Twitter compared to dating apps or private messages.

We were also able to detect behavioral differences between groups of people with different survey responses. For example, methamphetamine users were more likely to use sex-related words in their messages. Earlier research has shown that methamphetamine users have more sexual partners [55] and may be engaged in more risky sexual behavior, such as condomless anal sex, although it is not clear whether the relationship between methamphetamine use and condomless anal sex is causal [56]. Methamphetamine users were also less likely to travel far from home. This may be related to lower

XSL•FO
RenderX

household income level [57], which could make traveling far unaffordable, or paranoia induced by methamphetamine use. They also used more affective and social words and fewer drive-related words, which may be partly related to the higher prevalence of co-occurring mental health problems [57]. These insights could be valuable in personalizing methamphetamine use harm reduction and treatment by crafting prevention messages that frame behavioral modification as an affective or social process, instead of, for example, focusing on motivation.

We also found that participants with greater than 6 sex partners were more active users of all types of social apps (messaging, social media, and dating). Earlier studies have shown that users of geosocial networking apps, such as Grindr, have more sexual partners in general [58,59], and SGM with more partners have larger social networks [60], which may explain the more frequent use of social apps. Being more social may similarly explain more time spent away from home and in many different locations. People with greater than 6 sexual partners also used more sex- and drug-related words. Substance use has previously been found to be associated with a larger number of sexual partners [61]. These findings point toward personalized prevention that leverages both social media platforms and geolocation data. Partnerships between public health and dating apps to date have been limited to advertising and the addition of profile features (eg, fields for PrEP use). Researchers and public health practitioners might explore partnerships that allow users to opt into health promotion campaigns that are personalized by app use. Sexual health and substance use harm reduction should be pushed in relation to users' geolocation.

Comparison With Previous Work

The closest work to ours identified HIV as well as amphetamine, methamphetamine, and tetrahydrocannabinol use from social media messaging data [29]. Our work differs from this by using a wider range of data sources collected through participants' mobile devices. For example, we used text typed in any mobile app and website which allowed us to identify sexual risk behaviors in traditional messaging apps and less common dating apps in addition to the most popular social media apps. In addition, we used location data, which allows us to analyze participants' daily movement patterns and their relation to risk behaviors. We also attempted to identify a wider range of risk behaviors, especially related to sexual health. The only shared prediction target between these 2 studies was methamphetamine use; the earlier paper was able to predict with an F_1-score of 0.85, which was very close to our result (0.83). In summary, our work is aligned and expands previous research on this topic.

Another similar study predicted alcohol, tobacco, prescription drug, and illegal drug use from Instagram data [28]. They were able to detect alcohol use with statistical significance, but they had less success in predicting other types of substance use. Our better prediction results may be attributed to having access to more personal messaging data, as many people may avoid discussing substance use on public platforms. This shows that choosing the appropriate data collection methods is very important for accurate results.

Other studies have implemented personalized MSM interventions using survey data [62], identified the efficacy of MSM-targeted mobile app interventions [63], or evaluated the feasibility and acceptability of mobile sensing among MSM [64-66]. However, these studies have not evaluated whether mobile sensing data can be used to inform and personalize interventions, nor have they incorporated a broader SGM population inclusive of transgender and gender-diverse individuals, which were the goals of our study.

Limitations

A limitation of our study was that we only included participants who had an Android smartphone. We chose to only include Android users because iPhones have more restrictions on what data can be collected, and therefore collecting text data would have been unfeasible based on the budget for this project. The demographic differences between Android and iPhone users have been previously described, with iPhone users more likely to be female, younger, more concerned about their smartphone as a "status object," and displaying lower levels of honesty and humility and higher levels of emotionality [67]. The restriction of our study population to Android users only may limit the generalizability of results from this formative study. Potential participants had to be excluded from this study either because of the challenges with iOS adaptation (n=324) or due to missing data from Android users (n=11), which may skew the demographics to some extent, again impacting generalizability. In addition, as the data was collected using personal devices, there were some interruptions in data collection. For example, some participants turned off or deleted the app during the study while others upgraded to a new phone without reinstalling the app. Some Android phones were found to have aggressive battery-saving functionality which occasionally turned off the data collection. To avoid data collection issues, we kept track of when each participant's device had last sent us data and contacted participants after three missing days to make sure data collection could be resumed.

Our machine-learning model was trained on English-language text data only, which limits its ability to accurately interpret text written in other languages or in culturally specific dialects and vernaculars. While the ability to speak and understand English was one of our eligibility criteria, we did not exclude participants who are multilingual (ie, speak one or more languages other than English), and the model may not accurately process multilingual input. This limitation highlights the need for our future work to incorporate multilingual natural language processing approaches to better reflect diversity of language use.

Another limitation was that the text data only included what the participants typed on their phones. This approach may miss the context of some messages, as the responses are not collected. It could be informative to know what content participants consumed online or what messages they received from others. In addition, participants might have messaged with people using multiple devices (eg, computer or tablet in addition to their phone), so our data collection approach might not have been able to track all social media usage and messaging for some participants.

Finally, some of the outcomes that we set out to predict were very infrequent, which made the task impossible. For example,

XSL•FO
RenderX

only two of our participants were in a substance use treatment program, which was not enough for training and evaluating a machine learning model. Therefore, we had to focus on questions that had a reasonable number of both positive and negative responses.

Conclusions

In this study, we have shown that certain types of substance use and sexual risk behaviors can be determined from data that is collected from smartphones passively. Next, we demonstrated these data to be highly predictive of self-reported methamphetamine use and having 6 or more sexual partners. If integrated into downstream eHealth/mHealth interventions, passive mobile-sensing could be used to personalize interventions for SGM, which may reduce the burden of participating in intervention programs, as the daily behaviors can be tracked with minimal effort from the participant. However, further work is still needed to evaluate the efficacy of interventions based on automatic behavior tracking.

Our future work will explore providing personalized interventions using predictive models to determine which types of interventions may be appropriate. We will, for example, investigate sending participants messages and resources that are delivered "just in time," such as providing information about PrEP to individuals who may be at elevated HIV risk based on their substance use or sexual behavior but who are not yet taking it. This work will include developing interpretation guidelines for both automated systems and health care providers who may use these predictions in clinical settings.

Acknowledgments

This research was supported through a grant from the National Institute on Drug Abuse (DP2DA049296; Holloway). CC is supported as a RISE Fellow by the National Institutes of Health (R25GM061222). Generative artificial intelligence was not used in any portion of the current manuscript.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the formative nature of the research, small sample size, and plethora of private text-based information that could increase risk of identifiability of study participants, but are available from the corresponding author on reasonable request.

Authors' Contributions

MB: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft. KK: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft. CH: Investigation, Data curation, Project administration, Writing – review & editing. CC: Investigation, Data curation, Formal analysis, Project administration, Writing – review & editing. ESCW: Investigation, Data curation, Writing – review & editing. CB: Investigation, Data curation, Writing – review & editing. AA: Investigation, Data curation, Writing – review & editing. EY: Writing – review & editing. IWH: Conceptualization, Funding acquisition, Supervision, Project administration, Writing – review & editing. MS: Conceptualization, Methodology, Supervision, Writing – review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Risk assessment survey and location-based features.
[DOCX File, 17 KB - ojphi_v17i1e68013_app1.docx]

References

1. Medley G, Lipari RN, Bose J, Cribb DS, Kroutil LA, McHenry G. Sexual orientation and estimates of adult substance use and mental health: Results from the 2015 National Survey on Drug Use and Health. NSDUH Data Review 2016;10:1-54 [FREE Full text]
2. Bourne A, Weatherburn P. Substance use among men who have sex with men: patterns, motivations, impacts and intervention development need. Sex Transm Infect 2017 Aug;93(5):342-346. [doi: 10.1136/sextrans-2016-052674] [Medline: 28400466]
3. HIV surveillance report. : Centers for Disease Control and Prevention; 2017 URL: https://stacks.cdc.gov/view/cdc/60911 [accessed 2025-08-08]
4. HIV.gov. Who Is at Risk for HIV?. URL: https://www.hiv.gov/hiv-basics/overview/about-hiv-and-aids/who-is-at-risk-for-hiv [accessed 2023-02-14]
5. Centers for Disease Control and Prevention. Diagnoses, deaths, and prevalence of HIV in the united states and 6 territories and freely associated states, 2022. : HIV Surveillance Report; 2024 URL: http://www.cdc.gov/hiv-data/nhss/hiv-diagnoses-deaths-prevalence.html [accessed 2024-10-15]

XSL•FO
RenderX

6. Parsons JT, Rendina HJ, Moody RL, Ventuneac A, Grov C. Syndemic production and sexual compulsivity/hypersexuality in highly sexually active gay and bisexual men: further evidence for a three group conceptualization. Arch Sex Behav 2015 Oct;44(7):1903-1913. [doi: 10.1007/s10508-015-0574-5] [Medline: 26081246]
7. Singer M, Clair S. Syndemics and public health: reconceptualizing disease in bio-social context. Med Anthropol Q 2003 Dec;17(4):423-441. [doi: 10.1525/maq.2003.17.4.423] [Medline: 14716917]
8. Brooks VR. Minority Stress and Lesbian Women: Lexington Books; 1981.
9. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697. [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
10. Gonzales G, Henning-Smith C. Health disparities by sexual orientation: Results and implications from the behavioral risk factor surveillance system. J Community Health 2017 Dec;42(6):1163-1172. [doi: 10.1007/s10900-017-0366-z] [Medline: 28466199]
11. Budge SL, Adelson JL, Howard KAS. Anxiety and depression in transgender individuals: the roles of transition status, loss, social support, and coping. J Consult Clin Psychol 2013 Jun;81(3):545-557. [doi: 10.1037/a0031774] [Medline: 23398495]
12. Felner JK, Wisdom JP, Williams T, et al. Stress, coping, and context: Examining substance use among LGBTQ young adults with probable substance use disorders. Psychiatr Serv 2020 Feb 1;71(2):112-120. [doi: 10.1176/appi.ps.201900029] [Medline: 31640522]
13. Knight R, Karamouzian M, Carson A, et al. Interventions to address substance use and sexual risk among gay, bisexual and other men who have sex with men who use methamphetamine: A systematic review. Drug Alcohol Depend 2019 Jan 1;194(410-429):410-429. [doi: 10.1016/j.drugalcdep.2018.09.023] [Medline: 30502543]
14. Meiksin R, Melendez-Torres GJ, Falconer J, Witzel TC, Weatherburn P, Bonell C. eHealth interventions to address sexual health, substance use, and mental health among men who have sex with men: Systematic review and synthesis of process evaluations. J Med Internet Res 2021 Apr 23;23(4):e22477. [doi: 10.2196/22477] [Medline: 33890855]
15. Flores JM, Santos GM, Makofane K, Arreola S, Ayala G. Availability and use of substance abuse treatment programs among substance-using men who have sex with men worldwide. Subst Use Misuse 2017 Apr 16;52(5):666-673. [doi: 10.1080/10826084.2016.1253744] [Medline: 28139146]
16. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2020 National Survey on Drug Use and Health. : Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2021 URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2025-04-29]
17. Cascalheira CJ, Helminen EC, Shaw TJ, Scheer JR. Structural determinants of tailored behavioral health services for sexual and gender minorities in the United States, 2010 to 2020: a panel analysis. BMC Public Health 2022 Oct 12;22(1):1908. [doi: 10.1186/s12889-022-14315-1] [Medline: 36224564]
18. Sullivan PS, Driggers R, Stekler JD, et al. Usability and acceptability of a mobile comprehensive HIV prevention app for men who have sex with men: A pilot study. JMIR Mhealth Uhealth 2017 Mar 9;5(3):e26. [doi: 10.2196/mhealth.7199] [Medline: 28279949]
19. Dillingham R, Ingersoll K, Flickinger TE, et al. PositiveLinks: A mobile health intervention for retention in HIV care and clinical outcomes with 12-month follow-up. AIDS Patient Care STDS 2018 Jun;32(6):241-250. [doi: 10.1089/apc.2017.0303] [Medline: 29851504]
20. Ingersoll KS, Dillingham RA, Hettema JE, et al. Pilot RCT of bidirectional text messaging for ART adherence among nonurban substance users with HIV. Health Psychol 2015 Dec;34S(0):1305-1315. [doi: 10.1037/hea0000295] [Medline: 26651472]
21. Swendeman D, Ramanathan N, Baetscher L, et al. Smartphone self-monitoring to support self-management among people living with HIV: perceived benefits and theory of change from a mixed-methods randomized pilot study. J Acquir Immune Defic Syndr 2015 May 1;69 Suppl 1(0 1):S80-S91. [doi: 10.1097/QAI.0000000000000570] [Medline: 25867783]
22. Young SD, Yu W, Wang W. Toward automating HIV identification: Machine learning for rapid identification of HIV-related social media data. J Acquir Immune Defic Syndr 2017 Feb 1;74(Suppl 2):S128-S131. [doi: 10.1097/QAI.0000000000001240] [Medline: 28079723]
23. Krakower DS, Gruber S, Hsu K, et al. Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a modelling study. Lancet HIV 2019 Oct;6(10):e696-e704. [doi: 10.1016/S2352-3018(19)30139-0] [Medline: 31285182]
24. Marcus JL, Hurley LB, Krakower DS, Alexeeff S, Silverberg MJ, Volk JE. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. Lancet HIV 2019 Oct;6(10):e688-e695. [doi: 10.1016/S2352-3018(19)30137-7] [Medline: 31285183]
25. Wray TB, Luo X, Ke J, Pérez AE, Carr DJ, Monti PM. Using smartphone survey data and machine learning to identify situational and contextual risk factors for HIV risk behavior among men who have sex with men who are not on PrEP. Prev Sci 2019 Aug;20(6):904-913. [doi: 10.1007/s11121-019-01019-z] [Medline: 31073817]

XSL•FO
RenderX

26. Jing Y, Hu Z, Fan P, et al. Analysis of substance use and its outcomes by machine learning I. Childhood evaluation of liability to substance use disorder. Drug Alcohol Depend 2020 Jan 1;206:107605. [doi: 10.1016/j.drugalcdep.2019.107605] [Medline: 31839402]
27. Ahn WY, Ramesh D, Moeller FG, Vassileva J. Utility of machine-learning approaches to identify behavioral markers for substance use disorders: Impulsivity dimensions as predictors of current cocaine dependence. Front Psychiatry 2016;7:34. [doi: 10.3389/fpsyt.2016.00034] [Medline: 27014100]
28. Hassanpour S, Tomita N, DeLise T, Crosier B, Marsch LA. Identifying substance use risk based on deep neural networks and Instagram social media data. Neuropsychopharmacology 2019 Feb;44(3):487-494. [doi: 10.1038/s41386-018-0247-x] [Medline: 30356094]
29. Ovalle A, Goldstein O, Kachuee M, et al. Leveraging social media activity and machine learning for HIV and substance abuse risk assessment: Development and validation study. J Med Internet Res 2021 Apr 26;23(4):e22042. [doi: 10.2196/22042] [Medline: 33900200]
30. Holloway IW, Wu ESC, Boka C, et al. Novel machine learning HIV intervention for sexual and gender minority young people who have sex with men (uTECH): Protocol for a randomized comparison trial. JMIR Res Protoc 2024 Aug 20;13:e58448. [doi: 10.2196/58448] [Medline: 39163591]
31. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2023-07-15]
32. Zoom Video Communications, Inc. URL: https://zoom.us/ [accessed 2023-07-15]
33. The Platform Pushing What's Possible. Android. URL: https://www.android.com/ [accessed 2022-10-05]
34. Ferreira D, Kostakos V, Dey AK. AWARE: Mobile context instrumentation framework. Front ICT 2015;2(6). [doi: 10.3389/fict.2015.00006]
35. Moshe I, Terhorst Y, Opoku Asare K, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247. [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
36. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: Exploratory study. JMIR Mhealth Uhealth 2021 Jul 12;9(7):e26540. [doi: 10.2196/26540] [Medline: 34255713]
37. Vega J. Monitoring parkinson's disease progression using behavioural inferences, mobile devices and web technologies. Presented at: Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion); Apr 11-15, 2016; Montréal, Québec, Canada p. 323-327. [doi: 10.1145/2872518.2888598]
38. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47. [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
39. Preexposure prophylaxis for the prevention of HIV infection in the United States—2017 Update: a clinical practice guideline. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2017.pdf [accessed 2024-07-23]
40. Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 1966 Feb;10(8):707-710.
41. Life4. Life4/Textdistance: Compute distance between sequences 30+ algorithms, pure Python implementation, common interface, optional external libs usage.: GitHub URL: https://github.com/life4/textdistance [accessed 2022-10-05]
42. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit: O'Reilly Media, Inc; 2009.
43. Welcome to LIWC-22. LIWC. URL: https://www.liwc.app/ [accessed 2022-10-05]
44. Zhang X, Ghorbani AA. An overview of online fake news: Characterization, detection, and discussion. Inf Process Manag 2020 Mar;57(2):102025. [doi: 10.1016/j.ipm.2019.03.004]
45. Li L, Zhang Q, Wang X, et al. Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo. IEEE Trans Comput Soc Syst 2020 Mar 20;7(2):556-562. [doi: 10.1109/TCSS.2020.2980007]
46. Rathje S, Van Bavel JJ, van der Linden S. Out-group animosity drives engagement on social media. Proc Natl Acad Sci U S A 2021 Jun 29;118(26):e2024292118. [doi: 10.1073/pnas.2024292118] [Medline: 34162706]
47. Yin D, Bond SD, Zhang H. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. MIS Q 2014 Jun 1;38(2):539-560. [doi: 10.2307/26634939]
48. Alvero AJ, Giebel S, Gebre-Medhin B, Antonio AL, Stevens ML, Domingue BW. Essay content and style are strongly related to household income and SAT scores: Evidence from 60,000 undergraduate applications. Sci Adv 2021 Oct 15;7(42):eabi9031. [doi: 10.1126/sciadv.abi9031] [Medline: 34644119]
49. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Presented at: Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT); 2019 p. 4171-4186. [doi: 10.18653/v1/N19-1423]
50. Barbieri F, Camacho-Collados J, Espinosa Anke L, Neves L. TweetEval: unified benchmark and comparative evaluation for tweet classification. 2020 Presented at: Findings of the Association for Computational Linguistics; Online p. 1644-1650 URL: https://aclanthology.org/2020.findings-emnlp.148/ [accessed 2025-08-08] [doi: 10.18653/v1/2020.findings-emnlp.148]

XSL•FO
RenderX

51. Yizong Cheng C. Mean shift, mode seeking, and clustering. IEEE Trans Pattern Anal Machine Intell 1995;17(8):790-799. [doi: 10.1109/34.400568]
52. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011 Nov 1;12:2825-2830. [doi: 10.5555/1953048.2078195]
53. Cox DR. The regression analysis of binary sequences. J R Stat Soc Ser B 1958 Jul 1;20(2):215-232. [doi: 10.1111/j.2517-6161.1958.tb00292.x]
54. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Statist 2001 Oct;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
55. Gibson DR, Leamon MH, Flynn N. Epidemiology and public health Consequences of methamphetamine use in California's Central Valley. J Psychoactive Drugs 2002;34(3):313-319. [doi: 10.1080/02791072.2002.10399969] [Medline: 12422943]
56. Bryant J, Hopwood M, Dowsett GW, et al. The rush to risk when interrogating the relationship between methamphetamine use and sexual practice among gay and bisexual men. Int J Drug Policy 2018 May;55:242-248. [doi: 10.1016/j.drugpo.2017.12.010] [Medline: 29279253]
57. Jones CM, Compton WM, Mustaquim D. Patterns and characteristics of methamphetamine use among adults - United States, 2015-2018. MMWR Morb Mortal Wkly Rep 2020 Mar 27;69(12):317-323 [FREE Full text] [doi: 10.15585/mmwr.mm6912a1] [Medline: 32214077]
58. Goedel WC, Duncan DT. Geosocial-networking app usage patterns of gay, bisexual, and other men who have sex with men: Survey among users of Grindr, a mobile dating app. JMIR Public Health Surveill 2015;1(1):e4. [doi: 10.2196/publichealth.4353] [Medline: 27227127]
59. Hoenigl M, Little SJ, Grelotti D, et al. Grindr users take more risks, but are more open to human immunodeficiency virus (HIV) pre-exposure prophylaxis: Could this dating app provide a platform for HIV prevention outreach? Clin Infect Dis 2020 Oct 23;71(7):e135-e140. [doi: 10.1093/cid/ciz1093]
60. Smith AMA, Grierson J, Wain D, Pitts M, Pattison P. Associations between the sexual behaviour of men who have sex with men and the structure and composition of their social networks. Sex Transm Infect 2004 Dec;80(6):455-458. [doi: 10.1136/sti.2004.010355] [Medline: 15572613]
61. Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Schootman M, Cottler LB, Bierut LJ. Number of sexual partners and associations with initiation and intensity of substance use. AIDS Behav 2011 May;15(4):869-874. [doi: 10.1007/s10461-010-9669-0] [Medline: 20107887]
62. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874. [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
63. Yan J, Zhang A, Zhou L, Huang Z, Zhang P, Yang G. Development and effectiveness of a mobile phone application conducting health behavioral intervention among men who have sex with men, a randomized controlled trial: study protocol. BMC Public Health 2017 Apr 24;17(1):355. [doi: 10.1186/s12889-017-4235-6] [Medline: 28438144]
64. Trang K, Le LX, Brown CA, et al. Feasibility, acceptability, and design of a mobile ecological momentary assessment for high-risk men who have sex with men in hanoi, vietnam: qualitative study. JMIR Mhealth Uhealth 2022;10. [doi: 10.2196/preprints.30360] [Medline: 35084340]
65. Duncan DT, Kapadia F, Regan SD, et al. Feasibility and acceptability of global positioning system (GPS) methods to study the spatial contexts of substance use and sexual risk behaviors among young men who have sex with men in New York City: A P18 cohort sub-study. PLoS ONE 2016;11(2):e0147520. [doi: 10.1371/journal.pone.0147520] [Medline: 26918766]
66. Duncan DT, Chaix B, Regan SD, et al. Collecting mobility data with GPS methods to understand the HIV environmental riskscape among young Black men who have sex with men: A multi-city feasibility study in the deep south. AIDS Behav 2018 Sep;22(9):3057-3070. [doi: 10.1007/s10461-018-2163-9] [Medline: 29797163]
67. Shaw H, Ellis DA, Kendrick LR, Ziegler F, Wiseman R. Predicting smartphone operating system from personality and individual differences. Cyberpsychol Behav Soc Netw 2016 Dec;19(12):727-732. [doi: 10.1089/cyber.2016.0324] [Medline: 27849366]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
HIPAA: Health Insurance Portability and Accountability Act
IDU: Intravenous Drug Use
LIWC: Linguistic Inquiry and Word Count
MSM: men who have sex with men
NLTK: Natural Language Toolkit
PrEP: pre-exposure prophylaxis
SGM: sexual and gender minority
STI: sexually transmitted infection

XSL•FO
RenderX

Edited by E Mensah; submitted 25.10.24; peer-reviewed by HT Hsu, S Cook; revised version received 23.05.25; accepted 20.06.25; published 12.08.25.

Please cite as:
Beikzadeh M, Holloway IW, Kärkkäinen K, Hong C, Cascalheira C, Wu ESC, Boka C, Avendaño AC, Yonko EA, Sarrafzadeh M Identifying Substance Use and High-Risk Sexual Behavior Among Sexual and Gender Minority Youth by Using Mobile Phone Data: Development and Validation Study
Online J Public Health Inform 2025;17:e68013
URL: https://ojphi.jmir.org/2025/1/e68013
doi:10.2196/68013

© Mehrab Beikzadeh, Ian W Holloway, Kimmo Kärkkäinen, Chenglin Hong, Cory Cascalheira, Elizabeth S C Wu, Callisto Boka, Alexandra C Avendaño, Elizabeth Ann Yonko, Majid Sarrafzadeh. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 12.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Usage of Telehealth and Telenutrition Services by Registered Dietitian Nutritionists: Cross-Sectional Study

Najlaa Mohammed Al-Mana[1,2], PhD; Suhair Abdalla Abdalla[1], PhD; Asrar Abdulrahim Qari[1], MS; Mohamed Eldigire Ahmed[3], PhD; Wejdan Saeed Alshehri[1], BSc; Lujain Salem Baabdullah[1], BSc

[1]Clinical Nutrition Department, College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences-Jeddah, P.O. Box 9515, Jeddah, Saudi Arabia

[2]King Abdullah International Medical Research Centre, Ministry of the National Guard Health Affairs, Jeddah, Saudi Arabia

[3]Department of Basic Sciences, College of Science and Health Professions, King Saud Bin Abdulaziz University for Health Sciences-Jeddah, Jeddah, Saudi Arabia

Corresponding Author:
Najlaa Mohammed Al-Mana, PhD
Clinical Nutrition Department, College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences-Jeddah, P.O. Box 9515, Jeddah, Saudi Arabia

Abstract

Background: The COVID-19 pandemic has boosted telehealth adoption among clinical nutritionists globally. However, there is a research gap in Saudi Arabia concerning telehealth's prevalence and effectiveness in dietetics practice.

Objective: This study aims to evaluate telehealth implementation during the pandemic in Saudi Arabia.

Methods: In this cross-sectional study, a web-based survey was used and distributed in several Saudi Arabian regions between December 2022 and May 2023. A convenience sample of 306 clinical registered dietitian nutritionists (RDNs) in public and private health care facilities who met the study's inclusion criteria was included in this study.

Results: During the COVID-19 pandemic, 56% (172/306) of RDNs used telehealth, showing significant differences in sociodemographics and telehealth knowledge at health care facilities (P=.04). Notable gender disparities were observed in years of experience, with 78% (61/78) of male dietitians working in public hospitals (P=.001 and P<.004). The main telehealth nutrition services provided included nutrition education (21%, 64/306), nutrition counseling (19%, 58/306), and nutrition monitoring (17%, 52/306). Telenutrition purposes primarily focused on nutrition education (21%), supporting weight and diet management (17.15%, 15%), and the management of chronic disease (14%, 43,306). Additionally, a smaller percentage of RDNs (8%-9%) used telehealth for the nutrition care process and health assessment, while no respondents reported using telehealth for sport nutrition services. Overall, 90% (275/306) of RDNs reported that they routinely incorporated telehealth into their practice. Common obstacles reported by RDNs using telehealth were internet connectivity issues (46%), difficulties in coordinating with patients (22%), and patient disengagement with a lack of motivation (13%).

Conclusions: Our findings underscore the increasing adoption of telehealth by RDNs during the pandemic, highlighting its crucial role in nutrition services. The study suggests that technology enhancements and training initiatives can improve telehealth effectiveness, highlighting the need for further research in this dynamic field.

(Online J Public Health Inform 2025;17:e80211) doi:10.2196/80211

KEYWORDS

telehealth; dietetics; registered dietitian nutritionist; RDNs; nutrition education; telenutrition; COVID-19; Saudi Arabia

Introduction

The integration of technology into health care has revolutionized how medical services are delivered and accessed. Telehealth, a key innovation in this domain, has gained significant traction over the past decade, driven by advances in digital technologies and the growing demand for accessible health care solutions [1]. Telehealth is broadly defined as the delivery of health care services and information through virtual and telecommunication technologies to facilitate long-distance clinical care [2]. Tools such as video conferencing, mobile applications, and electronic communication platforms exemplify the diverse applications of telehealth [3].

Numerous studies have highlighted the advantages of telehealth, including reduced hospitalization rates, shorter lengths of stay, and enhanced continuity of care [4]. Telehealth has also played a pivotal role in protecting patients and health care providers by minimizing physical contact during infectious disease outbreaks [5]. Beyond its safety benefits, telehealth addresses critical barriers in health care delivery by reducing financial

XSL•FO
RenderX

burdens, overcoming geographical constraints, and fostering better usage of patient resources. Additionally, telehealth facilitates family involvement in care, even when family members are geographically distant [6].

Initially, telehealth adoption was limited primarily to primary care physicians. Over time, its application has expanded to include registered dietitian nutritionists (RDNs) and other health care professionals from various fields [7]. Dietitians can leverage telehealth for various purposes, including managing electronic health records, conducting consultations, and automating calculations such as BMI, calorie needs, and macronutrient distribution. Additionally, telehealth supports dietetic care beyond consultations by motivating patients, offering feedback, and enabling remote dietary monitoring [8]. The effectiveness of telehealth in dietetic care can be assessed based on various factors, including impact on dietitians' workflows, patient satisfaction, utility, and time efficiency [9]. A survey conducted among members of British, Australian, and New Zealand dietetic associations revealed that 62% of dietitians used telehealth for nutrition-related applications, of whom 74% used them as information resources and 60% used them for patient self-monitoring [10]. Recent findings highlighted a notable increase in dietitians using social and mass media platforms for telenutrition, from 68.9% before the COVID-19 pandemic to 80% during it. Additionally, 20% more dietitians became active in sharing nutrition information online, with 40% spending over 4 hours daily on these platforms as part of their practice during the pandemic [11].

Evidence from randomized controlled trials has further underscored the potential of telehealth in nutritional counseling. For example, telehealth consultations have successfully enhanced nutritional awareness and improved dietary practices among patients with age-related macular degeneration [12]. The value of telehealth becomes particularly evident during infectious disease outbreaks. In the United States, the proportion of clinical nutritionists providing care through telehealth increased from 37% before the COVID-19 pandemic to 78% during it [13].

Research conducted in Lima, Peru, demonstrated that virtual consultations achieved outcomes comparable to in-person visits regarding anthropometric changes in overweight and obese adults. This finding supports the viability of telehealth as an alternative when necessary [14]. Telehealth has proven particularly valuable during infectious disease outbreaks by minimizing physical contact while maintaining the quality of care. In Saudi Arabia, studies involving neurologists and neurosurgeons reported high satisfaction with telehealth services during the COVID-19 pandemic. These health care professionals highlighted telehealth's ability to foster stronger patient-provider relationships and enhance overall productivity. However, a notable challenge was the lack of physical examinations, which limited telehealth's suitability for follow-up visits rather than initial consultations [15]. A study examining RDNs' perspectives on telehealth found that 73.4% expressed optimism about its future role in dietetic practice [16]. However, further research is required to explore patients' experiences with telehealth and identify obstacles and unmet needs [17].

In Saudi Arabia, the adoption of telehealth has mirrored global trends, with significant progress in recent years [18]. However, the vast geographical expanse of the Kingdom and the concentration of specialized health care services in urban centers present significant barriers to equitable access, particularly for nutrition services in rural and remote areas [19]. Before the pandemic, nutrition service coverage faced challenges, with studies indicating issues such as low dietary diversity intake [20], and varying levels of nutrition care services for conditions such as obesity [21]. These preexisting challenges underscore a substantial local need for accessible nutrition interventions. Despite these needs, there remains a notable gap in research examining the prevalence and effectiveness of telehealth in dietetic practice within the Kingdom. To bridge this critical gap and address the urgent need for accessible nutrition care, this study aimed to evaluate the application of telehealth in dietetic practice during the COVID-19 pandemic in Saudi Arabia. This research sought to provide valuable insights into how telehealth was adopted, the types of services delivered, and the challenges faced by dietitians, informing future strategies to optimize virtual nutrition care and enhance service accessibility across the country.

Methods

Study Design and Participants

To evaluate the use of telehealth in dietetics practice during the COVID-19 pandemic in Saudi Arabia, we conducted an anonymous, web-based cross-sectional survey between December 2022 and May 2023. The study was implemented across several regions of Saudi Arabia and targeted RDNs working in both public and private hospitals and clinics.

Convenience sampling was used. The survey was distributed primarily through professional groups of dietitians on WhatsApp and Telegram, as well as through hospital and clinic administration by sending the survey link to heads of clinical nutrition departments, who then disseminated it to eligible staff. Additionally, outreach was conducted via targeted emails to RDNs and relevant professional associations to maximize participant reach.

A total of 306 participants aged 24 to 59 years were included, consisting of 257 females (84.0%) and 49 males (16%). Inclusion criteria were RDNs with a minimum of 1 year of professional experience who had used telehealth (virtual) as well as in-person visits in their practice. Exclusion criteria were dietitians with less than 1 year of experience, those who had not used telehealth, academics, and students in clinical nutrition departments.

Ethical Considerations

The study received ethical approval from the Ethics Research Committee at King Saud bin Abdulaziz University for Health Sciences (H-01-R-005). All participants were informed of the study's purpose, provided a declaration of consent, and gave informed consent before completing the survey. Participation was entirely voluntary, and no compensation, incentives, or reimbursements were offered. All procedures were conducted

XSL•FO
RenderX

in compliance with the ethical guidelines and regulations established by Saudi authorities.

Survey Questionnaire

The survey was developed using Google Forms and was available in both Arabic and English. The questionnaire was adapted from a previously validated instrument [16] and modified to suit the local context. To ensure content validity, the survey was reviewed by clinical nutrition experts and refined based on feedback from a pilot test conducted with a group of dietitians similar in demographic profile to the target population. Items were revised for clarity and relevance as necessary. The reliability of the study scale was assessed by Cronbach α coefficient for internal consistency, and the value was calculated at .74 showing a good level of internal consistency among scale items.

The questionnaire included sections on demographics (eg, age, gender, nationality, location, education, employment status, and patient population served), experiences with telehealth, perceived effectiveness, and barriers encountered during telehealth use.

Data Collection, Storage, and Security

Data were collected via secure, password-protected Google Forms. The responses were stored in encrypted digital files accessible only to the research team. All data collection and storage procedures were designed to maintain participant confidentiality and were conducted in accordance with institutional and national data protection policies. Only aggregated, anonymized data were used for analysis, and no personally identifiable information was collected. All research activities adhered strictly to the applicable ethical and data security guidelines.

Statistical Analysis

All collected data were cleaned and exported to IBM SPSS Statistics software (version 24.0) for analysis. A descriptive analysis was generated for all variables, with the quantitative variables being presented as mean and SD, and qualitative data reported in terms of frequency and percentage. Associations were tested using the χ^2, Fisher exact test, and Kruskal-Wallis test depending on the nature of the variable. A *P* value of .05 or lower was considered significant for all analytical tests.

Results

Overview

A total of 481 participants responded to the telehealth in dietetic practice survey, and 175 participants were excluded due to missing data, resulting in a total sample size of 306 participating RDNs who reported they have worked in both public and private health care institutions, had been in practice for over a year, and had experiences in both virtual and in-person visits were included in the study.

Sociodemographic Characteristics and Use of Telehealth at the Facility

Participant demographics, sociodemographic, and profession-related characteristics of dietitians working in hospitals and clinics in Saudi Arabia are presented in Table 1. Participant age ranged from 24 to 59 years, with 90% of them within the 24 to 34 years age range; the majority of them were female (257/306, 84%) with more than 84.3% of the RDNs holding a bachelor's degree (n=258).

XSL•FO
RenderX

Table . Sociodemographic characteristics of RDNs and use of telehealth among study sample (N=306).

Sociodemographic characteristics	Sex, n (%)		Fisher exact test	Chi-square *(df)*	*P* value	Overall
	Female (n=257)	Male (n=49)				
Age (y), n (%)			4.94[a]	—[b]	.089	
24 - 34	236 (91.8)	40 (81.6)				276 (90.2)
35 - 44	16 (6.2)	7 (14.2)				23 (7.5)
45 - 54	3 (1.2)	1 (2.0)				4 (1.3)
55 - 59	2 (0.8)	1 (2.0)				3 (1.0)
Nationality, n (%)			1.55[a]	—	.30	
Non-Saudi	17 (6.6)	1 (2.0)				18 (5.9)
Saudi	240 (93.3)	48 (98)				288 (94.1)
Province, n (%)			—	32.95 (4)[c]	<.001	
Central Province	102 (39.6)	22 (45)				124 (40.5)
Eastern Province	46 (18)	4 (8.2)				50 (16.3)
North Province	7 (2.7)	3 (6.1)				10 (3.3)
South Province	8 (3.1)	11 (22.4)				19 (6.2)
Western Province	94 (36.5)	9 (18.3)				103 (33.7)
Highest level of education, n (%)			4.44[a]	—	.108	
Bachelor's Degree	220 (85.6)	38 (77.6)				258 (84.3)
Doctorate	4 (1.6%)	3 (6.1)				7 (2.3)
Master's degree	33 (12.8%)	8 (16.3)				41 (13.4)
Employment setting, n (%)			—	11.0 (2)[c]	.004	
Other	87 (33.8)	8 (16.3)				95 (31.0)
Private	36 (14)	3 (6.1)				39 (12.7)
Public	133 (51.7)	38 (77.6)				171 (55.9)
Years of experience, n (%)			25.47[a]	—	<.001	
<1	79 (30.7)	2 (4.1)				81 (26.5)
1 - 5	132 (51.3)	30 (61.2)				162 (52.9)
5 - 10	28 (10.8)	10 (20.4)				38 (12.4)
10 - 15	11 (4.3)	1 (2.0)				12 (3.9)
15 - 20	3 (1.2)	4 (8.2)				7 (2.3)
>20	4 (1.6)	2 (4.1)				6 (2.0)
Is telehealth implemented at your institution during the pandemic? n (%)			—	15.1 (2)[c]	<.001	
Don't know	79 (30.7)	2 (4.1)				81 (26.5)
No	44 (17.1)	11 (22.4)				55 (18.0)
Yes	134 (52.1)	36 (73.4)				170 (55.6)

[a]Fisher exact tests were conducted for age, educational level, nationality, and years of experience.

[b]not applicable.

[c]Chi-square tests for association were conducted for province, employment setting, and telehealth implementation

XSL•FO
RenderX

The distribution of RDNs across different provinces was as follows: (124/306, 40.5%) in the central province, (103/306, 33.7%) in the western region, (50/306, 16.3%) in the eastern province, (19/306, 6.2%) in the southern province, and (10/306, 3.3%) in the northern province. The practice settings using telehealth comprise a mix of private hospitals, specialist clinics, research centers, and government hospitals caring for all age groups. Moreover, 56% of the respondents work as RDNs in public hospitals, whereas approximately 39% work at private hospitals and clinics. In comparison to other groups, RDNs stated that they work with adults the most frequently (40%), and approximately half of RDNs had 1 to 5 years of dietetics practice, in terms of experience.

There were significant variations in years of experience among dietitians, with notable differences observed between genders. Female dietitians tended to have less experience, with 30.7% (79/257) having less than 1 year of experience and 51.3% (132/207) having 1 to 5 years of experience. In contrast, male dietitians showed a different distribution, with 20.4% (10/49) having 5 to 10 years of experience and a higher percentage having over 20 years of experience at a senior level. This disparity in experience levels between male and female dietitians was statistically significant (*P*=.001).

The study highlighted gender disparities in the implementation of telehealth during the pandemic. A higher proportion of male dietitians (73%, 36/49) used telehealth compared to female dietitians (52%, 134/257). Additionally, male dietitians were more prevalent in public hospitals, with 78% (38/49) of male dietitians working in this setting (*P*<.01).

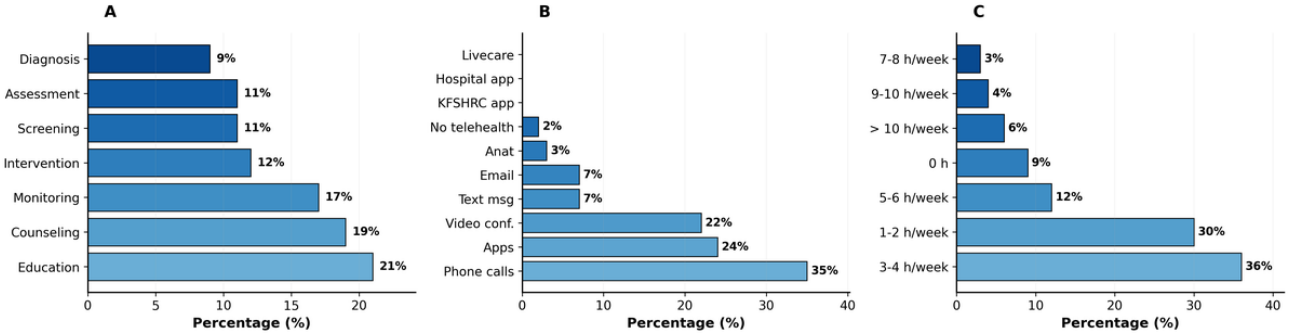
Telehealth Practice and Experiences During the COVID-19 Pandemic

During the COVID-19 pandemic, only 55.6% (170/306) of the RDNs confirmed they used telehealth at the clinic or hospitals, 18.0% (55/306) did not use any telehealth, whereas 26.5% (81/306) did not know whether telehealth was used or not. χ^2 tests revealed a significant difference across the sociodemographic domains and knowledge of the use of telehealth at the health facilities (*P*=.04; Table 1).

Nutrition Care Services Through Telehealth

According to Figure 1 (panel A), the majority of nutrition care services provided via telehealth during the COVID-19 pandemic included nutrition education, nutrition counseling, and nutrition monitoring. Respondents reported that typical aspects of the nutrition care process conducted through telehealth were nutrition education (21%, 64/306), nutrition counseling (19%, 58/306), and nutrition monitoring (17%, 52/306), with only 9% of study dietitians using telehealth for nutrition diagnosis.

Figure 1. Telehealth services utilization during COVID-19. (A) Nutrition services offered, (B) types of applications used, and (C) frequency of use in clinical practice. Data shown as percentages (N=306).



Our results identified the most common telehealth platforms used by RDNs, including phones and video conferencing platforms such as Zoom and MS Teams. Among the respondents, 35%(107/306) of RDNs reported providing nutrition intervention via telehealth through phone calls, followed by phone apps (eg, WhatsApp and Telegram) at 24% (73/306), and video conferencing (eg, Zoom and MS Teams) at 22% (67/306). Furthermore, only 7% (21/306) of RDNs indicated using emails or text messaging for nutrition care services, whereas a mere 3% (9/306) of RDNs used specialized applications such as Anat, a digital platform for health practitioners. Minimal usage (<1%) was observed for other institution-specific applications (Figure 1; panel B).

Purpose and Frequency of Telehealth Implementation

When asked about the purposes of implementing telehealth, the study participants highlighted that the leading objectives were nutrition education 21% (64/306) and weight management 17.15% (53/306). Additionally, diet management at 15% (46/306) and the management of chronic disease at 14% (43/306) were noted as key purposes, with 8% to 9% (24-28/306) focusing on the nutrition care process and health assessment, and 5% (15/306) for administrative use. Interestingly, no respondents reported using telehealth for sports nutrition purposes. Furthermore, a significant proportion of RDNs mentioned that they typically used telehealth with their patients 90% (275/306). The survey indicated that more than half of the RDNs use telehealth services weekly, with a notable majority of 36% (110/306) of them dedicating an average of 3 to 4 hours per week to providing nutrition care services remotely (Figure 1; panel C).

Obstacles and Challenges of Telehealth for RDNs

When using telehealth, RDNs encounter various obstacles and challenges across different employment settings. As per the findings presented in Table 2, the most prevalent obstacles experienced by RDNs during the implementation of telehealth were internet connection problems, affecting 46% (153/333) of participants. Interestingly, public settings reported higher rates of internet issues at 51% (78/170), with private settings close

XSL•FO
RenderX

behind at 44%(17/39). Another significant challenge was coordination issues, with approximately 22% (73/333) of respondents facing difficulties in coordinating with patients during telehealth sessions. Public settings showed a slightly higher prevalence of coordination challenges at 25% (43/170), similar to other settings at 25% (24/95), whereas private settings experienced a lower rate at 4%(2/39). Approximately 13% (43/333) of RDNs reported patient disengagement and lack of motivation as a challenge. Notably, public settings had slightly higher instances of this issue at 11% (19/170) compared to private settings at 7% (3/39), whereas other employment settings reported 25% (24/95). Both patients and providers encountered technology-related challenges, with 6% (20/333) reporting patient technology issues and 6% (20/333) reporting provider technology issues. Non–internet-related technology issues were more pronounced in private settings, with 11% (4/39)and 4% (2/39) reported, compared to public settings at 5% (9/170) and 1% (2/170), and other settings at 8% (8/95) and 0%.

Table . Obstacles and challenges faced by registered dietitian nutritionists in using telehealth services.

Challenges faced when using telehealth services	Overall (n=333), n (%)	By employment setting, n (%)			*P* value[a]
		Public	Private	Other	
Internet connection issues	153 (46)	87 (51)	17 (44)	24 (25)	.20
Patient disengagement and lack of motivation	43 (13)	19 (11)	3 (7)	24 (25)	.20
Inability to coordinate with a patient	73 (22)	43 (25)	2 (4)	24 (25)	.20
Patient technology issues (non–internet related)	20 (6)	9 (5)	4 (11)	8 (8)	.20
Provider technology issues (non–internet related)	20 (6)	2 (1)	2 (4)	0 (0)	.20
Others[b]	40 (12)	12 (7)	12 (30)	16 (17)	.20

[a]Kruskal-Wallis rank sum test.

[b]Other: physical examination issues, patient focus challenges, missed calls and complaints, patient commitment, lack of time and focus, patient engagement variations, technical education needs, and information understanding.

The “Others” category, representing 12% (40/333) of challenges and obstacles encountered by RDNs, included a diverse range of issues such as physical examination limitations, patient focus difficulties, missed calls impacting scheduling, patient commitment affecting appointments, knowledge gaps in tube feeding, time and focus constraints, varied patient engagement levels, technical education needs, and challenges in understanding information accurately. These challenges were most frequently reported among respondents from private employment settings at 30% (12/39), compared to 7% (12/170) in public settings and 17% (16/95) in other settings.

Discussion

Principal Findings

The COVID-19 pandemic had a profound impact on health care delivery worldwide, accelerating the adoption of telehealth as a critical tool for patient care. This rapid shift toward telehealth prompted health care providers and patients to adapt to virtual consultations, remote monitoring, and other telehealth services. This study examines telehealth usage among RDNs, focusing on the mode and frequency of use, as well as challenges of online health care navigation via telehealth as a care modality. The broader context of telehealth adoption in emerging economies, as highlighted by studies, such as Assaad et al, provides a valuable framework for understanding the unique dynamics at play in the Saudi Arabian health care landscape [22].

Saudi Arabia’s health care system is characterized by a predominantly public-funded and public-owned structure, providing free universal health care to its citizens. While there is a growing private sector, the government remains the largest operator, with ongoing reforms under Vision 2030 aiming to enhance efficiency and introduce privatization initiatives [23]. This unique context significantly influences health care delivery and the adoption of new technologies such as telehealth. Telehealth practice among RDNs during the COVID-19 pandemic showed various patterns [23]. Telehealth practice among RDNs during the COVID-19 pandemic showed various patterns [24]. The study results revealed that more than half of the RDNs had used telehealth in clinics or hospitals, whereas 18% reported not using any telehealth services, and 26% were uncertain about whether telehealth was implemented at their health facilities. Interestingly, the variability in telehealth adoption among RDNs was significantly influenced by sociodemographic factors (P=.04).

Concerning the cross-sectional study on the use of telehealth in the COVID-19 pandemic, a gender disparity existed (P<.001). A greater proportion of male participants (73%) reported that telehealth was adopted by their institution, compared to female participants (52%). Additionally, 31% of females were unaware of telehealth implementation, compared to 4.1% of male

XSL•FO
RenderX

participants. This knowledge gap indicates that male participants were more informed about the use of telehealth at their facilities. That said, because the vast majority of the sample was female participants, these results should be interpreted with caution. The gendered division of labor in Saudi health care is evolving, with increasing female participation, although disparities in leadership roles persist, influenced by societal norms [25,26]. This broader context of gender dynamics within the Saudi healthcare workforce may contribute to the observed differences in awareness and adoption of telehealth among male and female RDNs.

The 56% usage rate indicates that barely over half of the RDNs included telehealth into their practice during the pandemic, suggesting moderate acceptance and integrating telehealth technologies within the dietetics profession. This finding aligns with a recent study across 10 Arab countries by Bookari et al [11], which investigated dietitians' perceptions and practices regarding social or mass media use during the transition to telenutrition. Furthermore, a regional cross-sectional study by Bookari et al [27] provided a snapshot of the experiences of dietitians during the COVID-19 crisis in 5 Arab countries, highlighting similar trends and challenges in telehealth adoption. While a similar level of telemedicine implementation was also observed among RDNs in Argentina during the COVID-19 pandemic [28], the Saudi Arabian context, with its public hospital dominance and ongoing health care reforms, presents unique challenges and opportunities for telehealth integration that warrant specific attention. In contrast, medical and mental health care providers rapidly adopted telemedicine to ensure that they continued to deliver care without violating social distancing protocols [29,30].

The 18% of RDNs who had never used telehealth suggest a subgroup within the profession that is inexperienced with or unwilling to integrate such technologies. This might be due to constraints in gaining access to technology, training deficiencies, or self-perceived inefficiency in providing nutrition counseling through technology. Expanding on these factors will be pertinent to fully identify the reasons for telehealth reluctance or inability [8]. These barriers may require more detailed training conducted through the telehealth platforms, the availability of relevant technological equipment, and proving the effectiveness of telehealth in the provision of dietetics.

In this study, we found that 26% of the RDNs were unable to accurately describe the extent of telehealth integration in their health care facilities. This highlights a significant issue in communication and information dissemination within the healthcare sector, suggesting a lack of awareness about telehealth services. To address this, it is crucial for decision-makers at different levels to enhance internal communication and ensure that all health care professionals are well-informed about telehealth options and protocols [31].

Telehealth use during the pandemic also varied significantly by age. The majority of RDNs who used telehealth were between the ages of 24 and 34 years, compared to RDNs between the ages of 35 and 59 years who reported minimal use of telehealth. This is consistent with previous research indicating that younger generations are more willing and likely to learn and adopt new technologies [32-34], together highlighting a generational divide in technology adoption.

The levels of telehealth knowledge as functions of age, years of experience, and perhaps geographic location indicate potential factors that may affect or mitigate telehealth implementation. Younger or less experienced RDNs may adapt more easily to telehealth technologies, whereas more seasoned professionals may face challenges in using these tools. Similarly, access to telehealth functioning may be more readily available for RDNs in urban areas compared with rural contexts [35]. To address these disparities, measures must be taken to address barriers and concerns specific to different demographic groups [36]. The results present substantial provincial variations, the workplace distribution, with marked gender disparities.

The range of services delivered through telehealth by RDNs during the COVID-19 pandemic highlights the versatility of telehealth in nutrition practice. The data indicate that nutrition education was the most common service provided through telehealth (21%), nutrition counseling (19%), and nutrition monitoring (17%). These findings are consistent with previous studies by Law et al [37] and Kelly et al [38], which have emphasized the efficacy of telehealth in providing essential nutrition services. Our findings regarding nutrition monitoring in telehealth parallel research by Gnagnarella et al [31], who found that RDNs can use telehealth to document patients' progress and adherence to prescribed diets. Moreover, telehealth has enabled continuous nutrition monitoring, as supported by the work of Krznarić et al [39], who found that remote monitoring tools have significantly enhanced patient adherence to nutrition plans and improved health outcomes. Emphasizing the importance of these services in addressing patients' dietary and health needs, particularly in challenging times such as the pandemic [40,41]. The results of this study, alongside the previous research, underscore the adaptability and potential of telehealth in delivering comprehensive nutrition care services, ensuring that patients receive uninterrupted and quality care during unprecedented times.

The study reflects a strategic use of telehealth platforms by RDNs to maintain patient engagement and deliver essential nutrition services during the pandemic. Communications for nutrition education and follow-up consultations were primarily conducted via phone calls (35%), followed by phone applications (34%) and video conferencing tools (22%). Despite difficulties such as connectivity problems or patients' nonengagement, RDNs mostly described their work as enjoyable and promising, perceiving telehealth as future-oriented clinical work. While text and email communication were less common (7%), indicating a preference for more interactive channels, other platforms and institution-specific apps had minimal usage (<1%). These results align with studies in the United States and China indicating up to 53% use of technologies for communication and patients' follow-up routines [42,43].

The focus on nutrition education and counseling services suggests that although telehealth was employed to uphold key interventions and serve patients' continuing requirements, it remained most significant for ensuring that patients received the dietary guidance they needed while physical access was

XSL•FO
RenderX

disallowed. This aligns with other research by Hammersley et al [44], who conducted a review on telehealth nutrition services, reporting that they mainly involved education and counseling to keep patients on track by following professional dietary advice. This is also supported by Birkhoff and Moriarty [45], who found that telehealth changed the patient experience in that education and support could always be conducted when face-to-face consultations were not feasible.

The data suggest that before the pandemic, telehealth was employed mainly for informative and counseling interventions, which are core components of nutrition management that may be implemented via distance. A review by Sanaeinasab et al [46] noted that nutrition education and counseling are the fundamental components of telehealth services. This supports the conception of the present study that these services are being prioritized to deliver dietary necessities to patients physically restricted from consultations. Together, these studies affirm the centrality of nutrition education and counseling in telehealth and support our conclusion that these services are essential to sustain patients' care during disruptions such as the pandemic.

As expected, nutrition education was the most provided service, followed by weight management, diet management, and chronic disease management. These services are crucial in closely monitoring patients' conditions from a distance. This emphasizes the necessity of accessible and convenient media, such as phone calls and applications, for patient interaction and further care [13]. Conversely, the delivery of more specialized services such as metabolic and sports nutrition was notably infrequent, constituting less than 1% of the services provided. This could be attributable to the complexity and specificity of these services, which often require detailed physical assessments and tailored interventions that are challenging to perform remotely. Therefore, although telehealth can be effective for general nutrition counseling, it may have limitations in addressing the nuanced needs of specialized nutrition fields. However, a recent study by Eid et al [47] demonstrates the potential of telenutrition weight loss programs supported by telemonitoring and telehealth coaching to achieve positive anthropometric and biochemical outcomes in overweight and obese adults. These findings underscore the importance of further developing telehealth capabilities to better accommodate a broader range of nutrition services, ensuring comprehensive care delivery in both general and specialized nutrition domains. Nevertheless, the inconsistencies of telenutrition services indicate that some domains require further enhancement and inclusion into overall telehealth systems for effective and adequate nutrition support [11].

The difficulties faced by RDNs varied across different employment settings, providing insights into their diverse needs and identifying key areas that require optimal support [5]. Several concerns were exhibited regarding the use of telehealth, which differed between public and private employees' experiences. Internet connection issues were observed by 46% of RDNs, with the highest proportion in public (51%) as opposed to private (44%) and other (25%) practice settings, although not statistically significant (P=.2). Other drawbacks were physical examination, patients' distractedness, missed appointments, and lack of patient-caretaker information, which were reported by 12% overall, with the "other" setting being the highest at 30%, public at 7%, and private at 17%. These findings underscore the importance of targeted approaches to address these issues that could enhance telehealth service experiences for various employment settings [15,48,49]. Ethics and privacy are critical in the field of telehealth. To address these issues that were not addressed in this study, the implementation of appropriate privacy and security measures that are designed to protect patient information during virtual consultation is a priority. Providing RDNs with training on ethical standards and privacy regulations related to telehealth will help to ensure the protection of patient privacy and confidentiality.

This study has notable strengths. To our knowledge, this study is the first to evaluate telenutrition use across various regions in Saudi Arabia. The survey's validation and reliability were ensured by experienced RDNs practicing telenutrition, along with input from a focus group with diverse backgrounds, which helped ensure that the questionnaire was relevant and comprehensive for the target populations in special practice areas. However, this study has some limitations that should be considered when using the findings to develop telehealth practices and policies. First, the data collection was based on self-reports, which may be prone to respondents' memory or reporting bias. Second, the study employed convenience sampling and distributed surveys primarily through social media and online channels. This approach, while efficient during the pandemic, may have introduced selection bias by favoring younger or more technologically adept participants, potentially underrepresenting older RDNs or those less comfortable with digital platforms. Third, the participation was voluntary, and population bias may thus be an issue. The cross-sectional nature of the research limits our ability to indicate any causality or changes over time. Additionally, some subgroup analyses were constrained by small sample sizes, limiting the generalizability of these results. Finally, the study did not explore relative issues related to particular barriers and facilitators of telehealth, reducing the possibility of understanding the difficulties and prospects of RDNs in various settings.

Future Prospects

The future of telehealth in dietetics practice holds promising opportunities for innovation and advancement. To enhance the integration of telehealth, future efforts may focus on tailored training development, acquisition of technologies, and the effective internal communication in health care organizations. By incorporating emerging technologies, such as artificial intelligence and wearable devices, we may be able to elevate the delivery of personalized nutrition care and monitoring. The use of telehealth platforms in virtual nutritional assessments and meal planning that are tailored to the needs of the individual can revolutionize dietetic practice. There is a growing demand for further research to investigate the long-term effectiveness and outcomes of telehealth interventions in nutrition care. Future studies could focus on optimizing telehealth workflows, developing telehealth-specific training programs for RDNs, and exploring telehealth services to reach underserved populations. By embracing these future directions and implementing technological advancements, telehealth has the potential to

XSL•FO
RenderX

reshape and elevate the quality standards of care in dietetics practice.

Conclusions

Our study findings provide valuable insights that significantly contribute to the broader field of digital health in nutrition. The high usage of telehealth by RDNs during the COVID-19 pandemic underscores the growing importance and acceptance of digital health technologies in delivering nutrition care. While many RDNs used telehealth services effectively for education and counseling, several challenges were identified, including internet connectivity problems (affecting 46% of the respondents) and patients' nonattendance.

To address these obstacles, it is recommended that health care institutions focus on strengthening digital infrastructure, particularly in areas with unreliable internet connections to ensure consistent and high quality of telehealth delivery. Additionally, developing standardized and clear protocols for managing virtual appointments, including effective patient reminders and follow-up systems to reduce no-show rates. The implementation of targeted training programs will not only enhance the technical skills of RDNs in using telehealth platforms but also equip them with strategies for troubleshooting network issues and maintaining patient engagement in virtual settings. The findings suggest a need for increased research and special interventions on evaluating the effectiveness of these interventions and on exploring innovative solutions to technological barriers in digital health. Addressing limitations and challenges will be a key to embracing telehealth for enhanced access to high-quality nutrition services for all individuals. Overall, these findings expand our understanding of the evolving landscape of digital health in nutrition and provide a foundation for further research and development in this rapidly advancing field.

Data Availability

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

Conflicts of Interest

None declared.

References

1. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
2. Shaw DK. Overview of telehealth and its application to cardiopulmonary physical therapy. Cardiopulm Phys Ther J 2009 Jun;20(2):13-18. [Medline: 20467533]
3. Peregrin T. Telehealth is transforming health care: what you need to know to practice telenutrition. J Acad Nutr Diet 2019 Nov;119(11):1916-1920. [doi: 10.1016/j.jand.2019.07.020] [Medline: 31655720]
4. Peters GM, Kooij L, Lenferink A, van Harten WH, Doggen CJM. The effect of telehealth on hospital services use: systematic review and meta-analysis. J Med Internet Res 2021 Sep 1;23(9):e25195. [doi: 10.2196/25195] [Medline: 34468324]
5. Mehta P, Stahl MG, Germone MM, et al. Telehealth and nutrition support during the COVID-19 pandemic. J Acad Nutr Diet 2020 Dec;120(12):1953-1957. [doi: 10.1016/j.jand.2020.07.013] [Medline: 32792328]
6. Kaufman-Shriqui V, Sherf-Dagan S, Boaz M, Birk R. Virtual nutrition consultation: what can we learn from the COVID-19 pandemic? Public Health Nutr 2021 Apr;24(5):1166-1173. [doi: 10.1017/S1368980021000148] [Medline: 33436134]
7. Costantino A, Roncoroni L, Noviello D, et al. Nutritional and gastroenterological monitoring of patients with celiac disease during COVID-19 pandemic: the emerging role of telemedicine and point-of-care gluten detection tests. Front Nutr 2021;8:622514. [doi: 10.3389/fnut.2021.622514] [Medline: 33928109]
8. Ab Hamid MR, Abg Amar DR, Awang Basry NAS, Radzuan NN, Aulia Dewi AD. Utilisation of health information technology among dietitians in the workplace: a qualitative study. JUMMEC 2023;sp2023(1):43-50. [doi: 10.22452/jummec.sp2023no2.4]
9. Brunton C, Arensberg MB, Drawert S, Badaracco C, Everett W, McCauley SM. Perspectives of registered dietitian nutritionists on adoption of telehealth for nutrition care during the COVID-19 pandemic. Healthcare (Basel) 2021 Feb 23;9(2):235. [doi: 10.3390/healthcare9020235] [Medline: 33672179]
10. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. J Hum Nutr Diet 2017 Aug;30(4):439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]
11. Bookari K, Arrish J, Alkhalaf MM, et al. Perspectives and practices of dietitians with regards to social/mass media use during the transitions from face-to-face to telenutrition in the time of COVID-19: a cross-sectional survey in 10 Arab countries. Front Public Health 2023;11:1151648. [doi: 10.3389/fpubh.2023.1151648] [Medline: 37234764]
12. Tang D, Mitchell P, Liew G, Burlutsky G, Flood VM, Gopinath B. Telephone-delivered dietary intervention in patients with age-related macular degeneration: 3-month post-intervention findings of a randomised controlled trial. Nutrients 2020 Oct 10;12(10):3083. [doi: 10.3390/nu12103083] [Medline: 33050401]

XSL•FO
RenderX

13. Rozga M, Handu D, Kelley K, et al. Telehealth during the COVID-19 pandemic: a cross-sectional survey of registered dietitian nutritionists. J Acad Nutr Diet 2021 Dec;121(12):2524-2535. [doi: 10.1016/j.jand.2021.01.009] [Medline: 33612436]
14. Castrillón Liñan C, Alvarez Mayorga JH, Lozada-Urbano M. The effects of telenutrition in overweight and obese adults in a nutritional center in Lima, Peru. F1000Res 2021;10:545. [doi: 10.12688/f1000research.53564.2] [Medline: 34950451]
15. Algarni SA, Alrashid MH, Aldayel MS, et al. Satisfaction of adult and pediatric neurologists and neurosurgeons using telehealth during the COVID-19 pandemic in Saudi Arabia: a cross-sectional study. Front Digit Health 2024;6:1195697. [doi: 10.3389/fdgth.2024.1195697] [Medline: 38419806]
16. Zozula J, Smizer P, Prelack K. Telehealth in dietetic practice: current use and considerations for future applications. Top Clin Nutr 2021;36(2):137-146. [doi: 10.1097/TIN.0000000000000241]
17. Singh RH, Pringle T, Kenneson A. The use of telemedicine and other strategies by registered dietitians for the medical nutrition therapy of patients with inherited metabolic disorders during the COVID-19 pandemic. Front Nutr 2021;8:637868. [doi: 10.3389/fnut.2021.637868] [Medline: 33987197]
18. Alghamdi SM, Alqahtani JS, Aldhahir AM. Current status of telehealth in Saudi Arabia during COVID-19. J Family Community Med 2020;27(3):208-211. [doi: 10.4103/jfcm.JFCM_295_20] [Medline: 33354152]
19. Alzghaibi H, Hutchings HA. Barriers to the implementation of large-scale electronic health record systems in primary healthcare centers: a mixed-methods study in Saudi Arabia. Front Med (Lausanne) 2025;12:1516714. [doi: 10.3389/fmed.2025.1516714] [Medline: 40303363]
20. Hoteit M, Mortada H, Al-Jawaldeh A, et al. Dietary diversity in the Eastern Mediterranean region before and during the COVID-19 pandemic: disparities, challenges, and mitigation measures. Front Nutr 2022;9:813154. [doi: 10.3389/fnut.2022.813154] [Medline: 35252299]
21. Aljaaly EA, Khalifa NA. Comparison of nutrition care services for adult obesity at primary care in two different periods in Saudi Arabia. BMC Prim Care 2023 Jul 21;24(1):153. [doi: 10.1186/s12875-023-02094-6] [Medline: 37479976]
22. Assaad M, Chamma N, Mateev M, Rizk R. Telehealth during and beyond the COVID-19 pandemic: evidence from licensed dietitians in an emerging economy. Nutrition. Preprint posted online on 2024. [doi: 10.1101/2024.09.19.24314030] [Medline: 38883777]
23. Gurajala S. Healthcare system in the Kingdom of Saudi Arabia: an expat doctor's perspective. Cureus 2023 May;15(5):e38806. [doi: 10.7759/cureus.38806] [Medline: 37303448]
24. Appleton R, Williams J, Vera San Juan N, et al. Implementation, adoption, and perceptions of telemental health during the COVID-19 pandemic: systematic review. J Med Internet Res 2021 Dec 9;23(12):e31746. [doi: 10.2196/31746] [Medline: 34709179]
25. Al Mutair A, Al-Ghuraibi M, Alabbasi Y, Alguthaib F, Woodman A, Elgamri A. Saudi women's leadership experiences in the healthcare sector: a qualitative study. PLoS ONE 2023;18(9):e0285187. [doi: 10.1371/journal.pone.0285187] [Medline: 37725601]
26. Alobaid AM, Gosling C, Mckenna L, Williams B. Gendered organizational theory and glass ceiling: application to female Saudi paramedics in the workplace. Saudi J Health Sci 2020;9(3):177. [doi: 10.4103/sjhs.sjhs_56_20]
27. Bookari K, Arrish J, Zaher S, et al. A snapshot of the experience of dietitians during the COVID-19 crisis in five Arab countries: findings from a regional cross-sectional study. Nutrients 2022 Nov 19;14(22):4904. [doi: 10.3390/nu14224904] [Medline: 36432590]
28. Schlottke JA, Alvarez-Risco A, Bertiche EL, et al. Acceptance factors of telemedicine in times of COVID-19: case Argentina. PHAR 2024;71:1-10. [doi: 10.3897/pharmacia.71.e113509]
29. Cormi C, Sanchez S, de l'Estoile V, et al. Telepsychiatry to provide mental health support to healthcare professionals during the COVID-19 crisis: a cross-sectional survey among 321 healthcare professionals in France. Int J Environ Res Public Health 2021 Sep 27;18(19):10146. [doi: 10.3390/ijerph181910146] [Medline: 34639447]
30. Feijt M, de Kort Y, Westerink J, Bierbooms J, Bongers I, IJsselsteijn W. Integrating technology in mental healthcare practice: a repeated cross-sectional survey study on professionals' adoption of digital mental health before and during COVID-19. Front Psychiatry 2022;13:1040023. [doi: 10.3389/fpsyt.2022.1040023] [Medline: 36874171]
31. Gnagnarella P, Ferro Y, Monge T, et al. Telenutrition: changes in professional practice and in the nutritional assessments of Italian dietitian nutritionists in the COVID-19 era. Nutrients 2022 Mar 24;14(7):1359. [doi: 10.3390/nu14071359] [Medline: 35405971]
32. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Inf Syst J 2018 Nov;28(6):995-1019 [FREE Full text] [doi: 10.1111/isj.12179]
33. Győrffy Z, Boros J, Döbrössy B, Girasek E. Older adults in the digital health era: insights on the digital health related knowledge, habits and attitudes of the 65 year and older population. BMC Geriatr 2023 Nov 27;23(1):779. [doi: 10.1186/s12877-023-04437-5] [Medline: 38012565]
34. Spann A, Stewart E. Barriers and facilitators of older people's mHealth usage: a qualitative review of older people's views. Human Technology 2018:264-296 [FREE Full text] [doi: 10.17011/ht/urn.201811224834]
35. Osman S, Churruca K, Ellis LA, Braithwaite J. Beyond the planned and expected: the unintended consequences of telehealth in rural and remote Australia through a complexity lens. Med J Aust 2024 Jun 3;220(10):496-498. [doi: 10.5694/mja2.52294] [Medline: 38703008]

XSL•FO
RenderX

36. Nelson D, Inghels M, Kenny A, et al. Mental health professionals and telehealth in a rural setting: a cross sectional survey. BMC Health Serv Res 2023 Feb 27;23(1):200. [doi: 10.1186/s12913-023-09083-6] [Medline: 36849933]
37. Law L, Kelly JT, Savill H, et al. Cost-effectiveness of telehealth-delivered diet and exercise interventions: a systematic review. J Telemed Telecare 2024 Apr;30(3):420-437. [doi: 10.1177/1357633X211070721] [Medline: 35108135]
38. Kelly JT, Reidlinger DP, Hoffmann TC, Campbell KL. Telehealth methods to deliver dietary interventions in adults with chronic disease: a systematic review and meta-analysis. Am J Clin Nutr 2016 Dec;104(6):1693-1702. [doi: 10.3945/ajcn.116.136333] [Medline: 27935523]
39. Krznarić Ž, Bender DV, Laviano A, et al. A simple remote nutritional screening tool and practical guidance for nutritional care in primary practice during the COVID-19 pandemic. Clin Nutr 2020 Jul;39(7):1983-1987. [doi: 10.1016/j.clnu.2020.05.006] [Medline: 32425292]
40. Mottern M, Kharmats A, Curran M, et al. Impact of the COVID-19 pandemic on dietary counseling session attendance and self-monitoring adherence Dur034 a behavioral weight loss intervention. Curr Dev Nutr 2022 Jun;6:220. [doi: 10.1093/cdn/nzac048.034]
41. Avula R, Nguyen PH, Ashok S, et al. Disruptions, restorations and adaptations to health and nutrition service delivery in multiple states across India over the course of the COVID-19 pandemic in 2020: an observational study. PLoS ONE 2022;17(7):e0269674. [doi: 10.1371/journal.pone.0269674] [Medline: 35895693]
42. Khoshrounejad F, Hamednia M, Mehrjerd A, et al. Telehealth-based services during the COVID-19 pandemic: a systematic review of features and challenges. Front Public Health 2021;9:711762. [doi: 10.3389/fpubh.2021.711762] [Medline: 34350154]
43. Mechanic OJ, Persaud Y, Kimball AB. Telehealth Systems: StatPearls Publishing; 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK459384/ [accessed 2025-11-19]
44. Hammersley V, Donaghy E, Parker R, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Sep;69(686):e595-e604. [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
45. Birkhoff SD, Moriarty H. Challenges in mobile health app research: strategies for interprofessional researchers. J Interprof Educ Pract 2020 Jun;19:100325. [doi: 10.1016/j.xjep.2020.100325]
46. Sanaeinasab H, Saffari M, Yazdanparast D, et al. Effects of a health education program to promote healthy lifestyle and glycemic control in patients with type 2 diabetes: a randomized controlled trial. Prim Care Diabetes 2021 Apr;15(2):275-282. [doi: 10.1016/j.pcd.2020.09.007] [Medline: 33055009]
47. Eid NMS, Al-Ofi EA, Enani S, et al. Effects of a telenutrition weight loss program supported with telemonitoring and telehealth coaching on anthropometric and biochemical measures in overweight and obese adults: a pilot randomized controlled trial. Healthcare (Basel) 2024 Nov 25;12(23):2362. [doi: 10.3390/healthcare12232362] [Medline: 39684984]
48. Ramaswamy A, Yu M, Drangsholt S, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 9;22(9):e20786. [doi: 10.2196/20786] [Medline: 32810841]
49. Gustke SS, Balch DC, West VL, Rogers LO. Patient satisfaction with telemedicine. Telemed J 2000 May;6(1):5-13. [doi: 10.1089/107830200311806]

Edited by E Mensah; submitted 06.07.25; peer-reviewed by P Babatuyi, SB Guo; revised version received 28.09.25; accepted 05.11.25; published 27.11.25.

Please cite as:
Al-Mana NM, Abdalla SA, Qari AA, Ahmed ME, Alshehri WS, Baabdullah LS
Usage of Telehealth and Telenutrition Services by Registered Dietitian Nutritionists: Cross-Sectional Study
Online J Public Health Inform 2025;17:e80211
URL: https://ojphi.jmir.org/2025/1/e80211
doi:10.2196/80211

© Najlaa Mohammed Al-Mana, Suhair Abdalla Abdalla, Asrar Abdulrahim Qari, Mohamed Eldigire Ahmed, Wejdan Saeed Alshehri, Lujain Salem Baabdullah. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 27.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

E-Screening for Prenatal Depression in Kampala, Uganda Using the Edinburgh Postnatal Depression Scale: Survey Results

Hasifah Kasujja Namatovu[1*], DPhil; Mark Abraham Magumba[1*], DPhil; Dickens Akena[2*], DPhil

[1]Department of Information Systems, College of Computing and Information Sciences, Makerere University Kampala, Makerere University, P.O Box 7062, Kampala, Uganda

[2]Department of Psychiatry, Makerere University College of Health Sciences, Kampala, Uganda

[*]all authors contributed equally

Corresponding Author:
Hasifah Kasujja Namatovu, DPhil
Department of Information Systems, College of Computing and Information Sciences, Makerere University Kampala, Makerere University, P.O Box 7062, Kampala, Uganda

Abstract

Background: Perinatal depression remains a substantial public health challenge, often overlooked or incorrectly diagnosed in numerous low-income nations.

Objective: The goal of this study was to establish statistical baselines for the prevalence of perinatal depression in Kampala and understand its relationship with key demographic variables.

Methods: We employed an Android-based implementation of the Edinburgh Postnatal Depression Scale (EPDS) to survey 12,913 women recruited from 7 government health facilities located in Kampala, Uganda. We used the standard EPDS cutoff, which classifies women with total scores above 13 as possibly depressed and those below 13 as not depressed. The χ^2 test of independence was used to determine the most influential categorical variables. We further analyzed the most influential categorical variable using odds ratios. For continuous variables such as age and the weeks of gestation, we performed a simple correlation analysis.

Results: We found that 21.5% (2783/12,913, 95% CI 20.8% - 22.3%) were possibly depressed. Respondents' relationship category was found to be the most influential variable (χ^2_1=806.9, P<.001; Cramer's V=0.25), indicating a small effect size. Among quantitative variables, we found a weak negative correlation between respondents' age and the total EPDS score (r=−0.11, P<.001). Similarly, a weak negative correlation was also observed between the total EPDS score and the number of previous children of the respondent (r=−0.07, P<.001). Moreover, a weak positive correlation was noted between weeks of gestation and the total EPDS score (r=0.02, P=.05)

Conclusions: This study shows that demographic factors such as spousal employment category, age, and relationship status have an influence on the respondents' EPDS scores. These variables may serve as proxies for latent factors such as financial stability and emotional support.

(Online J Public Health Inform 2025;17:e51602) doi:10.2196/51602

KEYWORDS

perinatal; prenatal; antenatal; antepartum; depression; Edinburgh Postnatal Depression Scale

Introduction

Background

New and expectant mothers face several unique health challenges related to the physical and psychological changes accompanying motherhood. However, in low-income settings where mothers grapple with meeting basic physical needs such as access to safe water, proper nutrition, health care facilities, and health workers, psychological needs may end up being neglected. These needs are important factors as they can impact the physical health of the mother and the development of the child [1], and have been reported to increase the risk of stress, anxiety, and depression [2]. For pregnant women, these needs range from receiving affection and support throughout pregnancy, especially from spouses or partners, friends, and family [2]. Listening to their concerns, sharing resourceful information, and offering assistance are additional psychological needs of a pregnant woman [3]. Unmet needs, along with poverty, lack of spousal or partner support, early and unplanned pregnancies, and physical or verbal abuse, have accelerated perinatal depression [4-7]. Perinatal depression, that is depression that occurs during pregnancy, around childbirth, or within the first year post partum, affects households worldwide.

XSL•FO
RenderX

It often co-occurs with other medical or mental health illnesses and frequently goes undetected and untreated [8,9]. A study conducted by Fisher et al [10] reported that approximately 16% and 20% of women in low- and middle-income countries experienced antenatal and postnatal depression, respectively, with variations across settings.

Major depressive disorder in mothers affects 6% to 17% of pregnancies worldwide and can lead to negative outcomes such as preterm delivery, low birth weight [11,12], poor cognitive outcomes, and psychiatric morbidity in childhood and adolescence [13]. Furthermore, it affects a mother's ability to manage her children's feeding practices, contributes to poor socio-emotional development, and increases the likelihood of disruptive behavior [14,15]. In severe antenatal cases, perinatal depression may lead to suicidal ideation if left untreated [7,16]. A recent Lancet series highlighted the increasing global burden of mental health disorders, including maternal depression [17]. Although it is the most commonly diagnosed complication of childbirth, there are widespread gaps in screening, detection, and treatment of women affected by this incapacitating condition [18].

In Uganda, the prevalence of postnatal depression has been estimated to be 27% among patients attending primary care clinics [19], and yet it remains largely neglected, similar to other psychosocial disorders. The diagnosis of depressive illness is challenging and is sometimes prone to misdiagnosis [20]. This is largely attributed to (1) a lack of experience and knowledge of the professionals involved, (2) the complexity of clinical presentation [21], and (3) not consistently applying the routine diagnostic criteria during the initial evaluation process [22]. In Uganda, there are no formalized structures for depression screening during antenatal and postnatal visits, leaving individuals to self-diagnose, recognize their need for specialized assistance, and independently locate therapists. This is a significant challenge, considering some symptoms of depression may be attributed to other ailments, suggesting that afflicted individuals may even fail to effectively communicate their situation.

The World Health Organization recommends the integration of perinatal mental health care into primary care. However, in Uganda, this integration is not possible due to the vertical approach to service design, wherein maternal and mental health services are provided separately [23]. The goal of this paper is to establish statistical baselines for the prevalence of prenatal depression in Kampala and understand its relationship with key demographic variables.

Related Work

There have been several previous studies on maternal depression in Uganda and other sub-Saharan countries using a wide array of tools. For instance, Muhwezi et al [24] validated the 4-item subjective well-being subscale derived from the COMPASS OP Treatment Assessment System [25] in Uganda. Spies et al [26] and Baggaley et al [27] employed the 6-item and 10-item Kessler Screening Scale for Psychological Distress in South Africa and Burkina Faso, respectively. Kaaya et al [28] used the Hopkins Symptom Checklist-25, derived from the 90-item Symptom Checklist [29] in Tanzania. Kushniruk et al [30] employed the Centre for Epidemiological Studies Depression scale [28] in both Tanzania and South Africa. Chibanda et al [31] validated the Edinburgh Postnatal Depression Scale (EPDS) [32] in Zimbabwe. Two broad approaches are usually employed for tool validation, namely a psychiatrist's evaluation or comparison with gold standard tools. All the above studies except that by Baggaley et al [27] employed the Mini-International Neuropsychiatric Interview tool for validation [33]. Baggaley et al [27], considered a psychiatrist's assessment as the gold standard. Other tools mentioned in the literature, although not necessarily used in sub-Saharan Africa include the 9-item Patient Health Questionnaire, 20-item Self-Report Questionnaire, and Visual Analogue Scale. However, the EPDS is more widely applied, particularly for maternal depression screening.

The vast majority of studies are geared toward tool validation. Although tool validation is a component of our study, the primary objective is to establish statistical baselines for antenatal maternal depression in Kampala, Uganda; therefore, we employed a comparatively large number of respondents. The data comprises 12,913 records of EPDS results collected from pregnant women attending antenatal clinics in 7 health facilities in Kampala namely: Kitebi Health Centre HCIV, Kawempe Mulago Referral Hospital, Kampala Capital City Authority (KCCA) Health Centre HCIII, Bugolobi Health Centre HCIV, Mengo Kisenyi Health Centre HCIV, Kasubi Kawaala Health Centre HCIII, Komamboga Health Centre HCIII, Kasangati Health Centre HCIII. The data was collected between January 2022 and April 2022.

Methods

Ethical Considerations

This study was approved by the Research Ethics Committee of Mulago Hospital (MHREC 2021-57) and the Uganda National Council for Science and Technology (SS945ES). Additional approvals were obtained from KCCA as required for research conducted in KCCA facilities (DPHE/KCCA/1301). Written consent was obtained from all participants for participation and publication of the findings. The data were deidentified. Each respondent was given financial compensation of USh 20,0000 (US $5)

Study Design

The study participants were expectant mothers receiving antenatal care. The data was collected by 17 research assistants using an Android-based implementation of the EPDS. The results were then transmitted to a remote MySQL server instance via an internet connection. The respondents were required to have signed a consent form. In addition to the 10-question EPDS tool, they completed a general questionnaire that collected basic demographic data and contact information for potential follow-up, for instance, during the validation phase. This questionnaire was also administered in an electronic form via the Android platform.

Study Setting

In Kampala, there are a total of 1458 health facilities, comprising 26 government-owned, 1371 private-for-profit, and 61

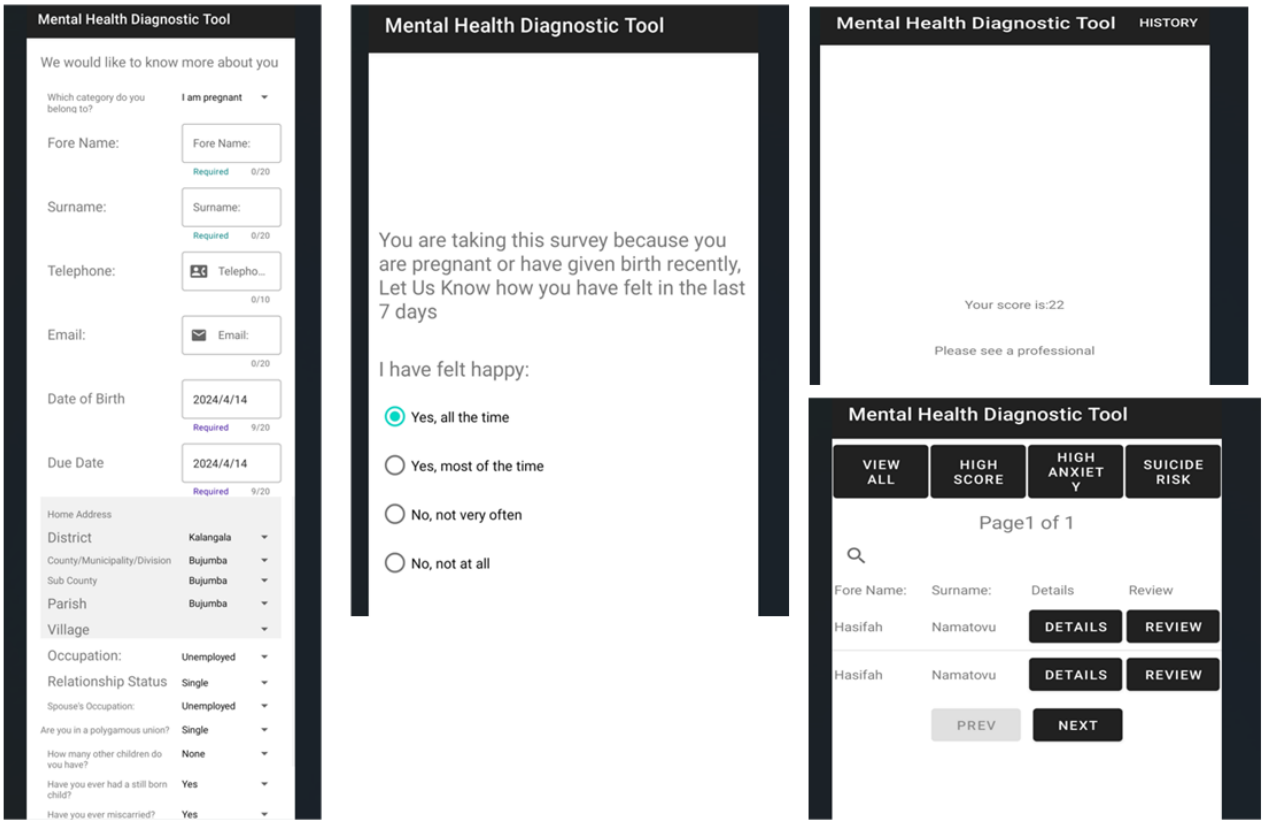
XSL•FO
RenderX

private-not-for-profit establishments [34]. Of these, 11 government-owned facilities provide antenatal care services to the public. In this study, we focused on 7 of the 26 government-owned facilities.

Data Collection Instrument (Android Implementation of EPDS)

The EPDS tool was automated using its standard 10 questions and was implemented via the Android platform installed on the tablet computers as shown in Figure 1. The responses for each of the 10 questions ranged from 0 to 3. For questions 3 and 5 - 10, the responses were organized in reverse order (3, 2, 1, and 0). The total score was obtained by adding up the individual scores for all 10 questions. The EPDS was developed to assist health care professionals in identifying mothers experiencing postnatal depression, although it has also been used to screen symptoms of depression in pregnant women [33]. For each of these 10 questions, the woman chose one of four responses reflecting how she felt a week prior to administration of the tool. The responses scored from 0 to 3 are based on the severity of the symptom, with 0 being less severe and 3 being more severe. For safety purposes, a woman who scored 1 or higher on question 10 (indicating suicidal ideation) was immediately referred for further diagnosis since the EPDS is only a screening tool, and not a diagnostic one. Similarly, women who scored 9 or higher were referred for follow-up, as advised by Cox et al [33]. The maximum possible score is 30, with a score of 10 and above indicating possible depression. Women who obtained a score above 13 could possibly be experiencing depressive illness that may vary in severity. To analyze the influence of specific factors on perinatal depression, demographic information was collected as part of the initial sections prior to the administration of the 10 questions of the EPDS.

Figure 1. Android implementation of the Edinburgh Postnatal Depression Scale.



Study Participants and Selection

A total of 12,913 prenatal women were recruited for the study from the 7 health facilities. Perinatal services were offered from Monday to Friday, from 8 AM to 1 PM at these facilities. During their visits, the research team was allocated a specific area where they could engage with expectant mothers for depression screening using the Android version of the EPDS. After completing the vital signs assessment, the research assistants randomly selected a respondent for screening. As the EPDS tool was not designed for self-administration, the research assistants would interact with women directly, following the sequence of questions in the scale. This approach was chosen as a significant number of the women were either illiterate or lacked the digital skills required to use a smartphone. To avoid the possibility of a woman discussing their responses, the screening was done in solitude.

Data Analysis

Data was analyzed using Microsoft Excel. We performed both univariate and bivariate analyses. For the univariate analysis, we established the relative proportions of respondents for each demographic variable. For the bivariate analysis, we examined the relationship between demographic variables and EPDS scores. For categorical variables, we relied on χ^2 tests of independence and risk and odds ratios. We employed a 2-step procedure; first, the χ^2 test of independence was used to determine the demographic variables with the strongest

XSL•FO
RenderX

association with the overall EPDS depression categories. Following this, we used the odds ratios to determine which specific values of these variables implied a higher risk for depression. For nonbinary variables such as the spousal or partner employment category, odds ratios were computed as the ratio of odds for respondents in that category to the odds for all other respondents for that variable in a "one versus all" setup. For continuous variables, including the weeks of pregnancy, respondents' age, and the number of previous children, a correlation analysis was conducted.

Inclusion Criteria

We included any patient of the antenatal clinics at the respective health facilities who were willing to complete the questionnaire and subsequent EPDS tool.

Results

Univariate Analysis

Approximately 66.9% (8633/12,913) of respondents reported being in polygamous unions, 29.7% (n=3838) were first-time mothers, and 49.8% (n=6431) reported having 2 or more previous children. Nearly 11% (1419) of respondents reported having a previously stillborn child, whereas 30.7% (3964) reported experiencing a miscarriage. The average number of children per woman was 1.6. The mothers' ages ranged from 9-53 (mean 26, SD 5.3) years. The mean EPDS score was 9.8 (SD 3.9), while the median score was 10 (IQR 4). We used the commonly accepted cutoff of 13 points or higher on the EPDS as an indicator of depression [35]. Consequently, we found that 2783 respondents (21.5%; 95% CI 20.8 - 22.3) were identified as possibly depressed. These findings have been presented in Table 1.

XSL•FO
RenderX

Table . Statistical summary of postnatal EPDS data from Kampala health facilities.

Demographics	EPDS[a] category			Likelihood/odds ratio (95% CI)	Pearson Chi-square		Cramer's V (effect size)
	Depressed, n (%)	Not depressed, n (%)	Total, n (%)		Chi square (*df*)	*P* value	
Education					76 (4)	<.001	0.08 (negligible)
A-level	300 (2.3)	1125 (8.7)	1425 (11)	1.0 (0.8 - 1.1)			
Degree/diploma	98 (0.8)	264 (2.0)	362 (2.8)	1.4 (1 - 1.7)			
Lower primary/none	289 (2.2)	707 (5.5)	996 (7.7)	1.5 (1.3 - 1.8)			
O-level	1534 (11.9)	6403 (49.6)	7937 (61.5)	0.7 (0.7 - 0.8)			
P5/higher	562 (4.3)	1631 (12.6)	2193 (17.0)	1.3 (1.2 - 1.5)			
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—[b]			
Respondent occupation					129.63 (4)	<.001	0.1 (small)
Formal business owner	200 (1.5)	730 (5.7)	930 (7.2)	1.0 (0.8 - 1.2)			
Formal employee	236 (1.8)	338 (2.6)	574 (4.4)	2.7 (2.3 - 3.2)			
Informal business owner	370 (2.9)	1808 (14.0)	2178 (16.8)	0.7 (0.6 - 0.8)			
Informal employee	223 (1.7)	595 (4.6)	818 (6.3)	1.4 (1.2 - 1.6)			
Unemployed	1754 (13.6)	6659 (51.6)	8413 (65.1)	0.9 (0.8 - 1.0)			
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—			
Spouse/partner's occupation					512 (4)	<.001	0.19 (small)
Formal business owner	243 (1.9)	1423 (11.0)	1666 (12.9)	0.6 (0.5 - 0.7)			
Formal employee	579 (4.5)	1611 (12.5)	2190 (16.9)	1.4 (1.2 - 1.5)			
Informal business owner	443 (3.4)	3132 (24.2)	3575 (27.5)	0.4 (0.4 - 0.5)			
Informal employee	829 (6.4)	2605 (20.1)	3434 (26.6)	1.2 (1.1 - 1.3)			
Unemployed	689 (5.3)	1359 (10.5)	2048 (15.8)	2.1 (1.9 - 2.4)			
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—			
Relationship status					806.9 (5)	<.001	0.25 (small)
Cohabiting	1644 (12.7)	4996 (38.7)	6640 (51.4)	1.5 (1.4 - 1.6)			
Formal union	157 (1.2)	530 (4.1)	687 (5.3)	1.1 (0.9 - 1.3)			
Separated	159 (1.2)	306 (2.4)	465 (3.6)	1.9 (1.6 - 2.4)			
Single	418 (3.2)	730 (5.7)	1148 (8.9)	2.3 (2.0 - 2.6)			
Traditional marriage	383 (3.0)	3540 (27.4)	3923 (30.4)	0.3 (0.26 - 0.33)			
Widowed	22 (0.2)	28 (0.2)	50 (0.4)	2.9 (1.6 - 5.0)			
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—			
Polygamous union					237.9 (1)	<.001	0.13 (small)
Yes	1506 (11.7)	7127 (55.2)	8633 (66.9)	0.5 (0.5 - 0.5)			
No	1277 (9.8)	3003 (23.2)	4280 (33.1)	2 (1.8 - 2.2)			

XSL•FO
RenderX

Demographics	EPDS[a] category			Likelihood/odds ratio (95% CI)	Pearson Chi-square		Cramer's V (effect size)
	Depressed, n (%)	Not depressed, n (%)	Total, n (%)		Chi square (*df*)	*P* value	
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—			
Previous miscarriage					211.6 (1)	<.001	0.12 (small)
Yes	1186 (9.2)	2782 (21.5)	3968 (30.7)	2 (1.8 - 2.1)			
No	1597 (12.3)	7348 (57.0)	8945 (69.3)	0.5 (0.5 - 0.6)			
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—			
History of stillbirth					2.99 (1)	<.001	0.01 (negligible)
Yes	345 (2.7)	1133 (8.7)	1478 (11.4)	1.1 (1.0 - 1.3)			
No	2438 (18.8)	8997 (69.7)	11,435 (88.6)	0.9 (0.8 - 1.0)			
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—			
Health facility					—	—	—
Kasangati HCIII	1 (0.0)	242 (1.9)	243 (1.9)	0.0 (0.0 - 0.1)			
Kasubi Kawaala HCI-II	304 (2.4)	1091 (8.4)	1395 (10.8)	1.0 (0.9 - 1.2)			
Kawempe Mulago Referral Hospital	465 (3.6)	2661 (20.6)	3126 (24.2)	0.6 (0.5 - 0.6)			
KCCA[c] HCIII - Bugolobi	842 (6.5)	538 (4.2)	1380 (10.7)	7.7 (6.9 - 8.7)			
Kitebi HCIV	562 (4.3)	646 (5.0)	1208 (9.4)	3.7 (3.3 - 4.2)			
Komamboga HCIII	257 (2.0)	3306 (25.6)	3563 (27.6)	0.2 (0.2 - 0.2)			
Mengo Kisenyi HCIV	352 (2.7)	1646 (12.7)	1998 (15.4)	0.7 (0.7 - 0.8)			
Total	2783 (21.5)	10,130 (78.4)	12,913 (100.0)	—			

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]Not applicable.

[c]KCCA: Kampala Capital City Authority.

Education Level

The most common education level was O-level, accounting for 61.2% (n=7937), followed by higher primary school (P5 or higher) at 17% (n=2193), A-level education at 11% (n=1425), and lower primary school at approximately 7.7% (n=996), as demonstrated in Table 1. Lastly, tertiary level education (degree or diploma), accounted for 2.8% (n=362). In Uganda, O-level corresponds to the Ordinary Level Certificate of Education, comparable to the UK's General Certificate of Education, whereas A-Level corresponds to the Advanced Certificate of Education.

Respondents' Employment Status

The most common employment category among respondents was unemployed, accounting for 65.1% (8413/12,913). This was followed by 16.8% (n=2178) informal business owners, 7.2% (n=930) formal business owners, 7.2% informal employees (n=930), and 4.4% (n=574) formal employees. Formal business owners refer to individuals who own formally registered businesses, whereas informal business owners include market vendors, hawkers, and small-scale farmers who own informal businesses. Informal employees are those engaged in short-term work on a daily or weekly basis, without a formal contract, whereas formal employees are employed on a long-term basis with well-stipulated contracts, similar to the majority of white-collar workers.

Employment Status of Respondents' Spouse or Partner

The most common employment category of respondents' spouses or partners was informal business owners, accounting for 27.5% (n=3575) of respondents, followed by those with informally employed spouses or partners representing 26.6% (n=3434) of respondents. Nearly 16.9% (n=2190) of respondents had spouses or partners who were formal employees, whereas 15.8% (n=2048) of respondents had unemployed spouses, and

XSL•FO
RenderX

12.9% (n=1666) of respondents had spouses who were formal business owners.

Relationship Status

The most common relationship status among respondents, accounting for 51.4% (n=6640) of respondents, was cohabiting representing partners living together with no formal or traditional marriage, followed by traditional marriages representing 30.4% (n=3923) of respondents. Single women comprised 8.9% (n=1148) of respondents, while those in formal unions (church, mosque, or civil weddings) comprised 5.3% (n=687) of respondents, and women who had separated from their spouses or partners constituted 3.6% (n=465) of respondents. Lastly, widowed women comprised 0.4% (n=50) of respondents.

Polygamous Unions

This variable refers to whether the respondent was in a relationship with a partner who had one or more additional sexual partners. A majority (n=8633, 66.9%) of respondents reported that they were in polygamous relationships.

Previous Miscarriage

This variable refers to whether a woman had previously experienced a miscarriage. We found that nearly 30.7% (n=3968) of respondents reported a history of miscarriage, in agreement with previously published studies [36].

Previous Stillbirth

This variable refers to whether the respondent had previously given birth to a stillborn child. We found that a total of 11.4% (n=1478) of respondents reported a stillborn child. These results align closely with previously reported data [37].

Bivariate Analysis

For the bivariate analysis, we initially used the χ^2 test to determine which demographic features had a strong influence on the EPDS classification. We then calculated the odds ratio for each feature to understand its effects better. Table 1 includes odds ratios for various health facilities for completeness; however, we excluded the hospital as a variable for determining depression risk. This decision was due to variable data collection periods across different facilities, indicating that the sample sizes may not accurately reflect the typical volume of clients for each facility, making cross-facility comparisons statistically unreliable.

Among the remaining categories (education level, respondents' occupation, spouse or partners' occupation, polygamous union, previous miscarriage, and stillbirth), the respondents' relationship status had the largest effect on their EPDS score (χ^2=806.9, P<.001; Cramer's V=0.25). The effect size was determined using Cramer's V and *df* as per the method described by Kim [38]. As seen in Table 1, widowed women had the highest odds ratio for depression and were nearly 3 times more likely to be depressed compared to nonwidowed women. Conversely, women in traditional marriages were the least likely to report depression compared to other groups.

For quantitative variables, we found a negative correlation between respondents' age and the total depression score (r=−0.11, P<.001). Similarly, there was a negative correlation between total depression and number of previous children (r=−0.11, P<.001). A weak negative correlation was also observed between weeks of gestation and the total depression score (r=−0.03, P=.002).

Discussion

Principal Findings

We found an overall prevalence of 21.5% (95% CI 20.8%-22.3%). We could not locate previous studies from Uganda that focused specifically on prenatal or antenatal mothers and used the EPDS; this prevalence falls within the range of previous studies in similar regions [39,40]. Relationship status was the most important determinant of depression. The group most likely to display depressive symptoms was widowed women. Broadly, women who referred to themselves as single, separated, or widowed displayed a significantly higher propensity for depression compared to those who were partnered. This outcome is in agreement with previous studies [41] and is not surprising given that spousal or partner support has been widely recognized as an important factor [40,42,43].

The correlation analysis revealed a weak negative relationship between maternal depression during pregnancy and the number of children. Although we found no previous studies specifically investigating this correlation, there is indirect corroboration from studies suggesting that women with more children reported lower levels of depression [44,45]. However, the direction of causality may be reversed, as individuals prone to depression are also less likely to have more children [46]. Furthermore, the correlation analysis revealed a weak negative association between age and antenatal depression, in agreement with some previous studies [47]. However, other studies have reported contrary findings [48]. This may be attributed to age being correlated with latent variables such as income and parental experience, which may vary across different cohorts. For instance, we found a very strong positive correlation between age and the number of children implying that one of these variables may be confounding. We also observed a negative correlation between the depression scores and weeks of gestation. This is in agreement with some previous studies that have noted a reduction in depression as pregnancy progresses [49]. However, the association in our study was too weak to merit extensive discussion.

Limitations of the Study

A key limitation of this study is its reliance on indirect parameters such as spousal or partner employment categories, rather than directly measuring household income and spousal support. This limitation arises because the demographic data were originally collected as preliminary input for creating user profiles in the EPDS Android app. Important missing parameters that could provide deeper insights into the woman's depressive state include social support, intimate partner violence, early and unplanned pregnancies, physical and verbal abuse, self-esteem issues, emotional abuse, and physical health parameters concerning the pregnancy and the woman's overall health. These

XSL•FO
RenderX

factors could be explored in follow-up studies with smaller, more focused cohorts. However, this study provides a valuable statistical overview of prenatal depression in Kampala, Uganda.

Implications of the Study

These findings imply that with sufficient data, empirical approaches could be developed to identify individuals at risk of depression. For instance, demographic markers such as respondents' relationship statuses have a strong influence on the EPDS depression categories and could be used to create risk profiles or automate interventions.

Conclusion

Perinatal screening for depression is not commonly performed during and after pregnancy, despite being emphasized by the World Health Organization. This study revealed that women whose spouses or partners were engaged in some form of employment, especially those in informal business ownership were less likely to experience depression.

Acknowledgments

We extend sincere gratitude to the management of Kampala Capital City Authority and all participating health facilities for their support during data collection. Similarly, appreciation goes out to the research assistants and prenatal and postnatal mothers whose participation was crucial for the success of the study. Finally, we express our sincere appreciation to the Government of Uganda for funding this research through the Research and Innovations Fund at Makerere University.

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee support for the publication of this article.

Authors' Contributions

Conceptualization: HKN, MAM
Data collection: HKN, MAM
Methodology: HKN, MAM
Validation: HKN, MAM
Data curation: HKN, MAM
Statistical analysis: HKN, MAM
Android and Web Development: MAM
Writing – original draft preparation: HKN, MAM
Writing – review and editing: HKN, MAM, DA
Supervision: HKN

Conflicts of Interest

None declared.

References

1. Poppert Cordts KM, Wilson AC, Riley AR. More than mental health: parent physical health and early childhood behavior problems. Physiol Behav 2017;176(12):139-148. [doi: 10.1097/DBP.0000000000000755]
2. All about providing emotional support to women during pregnancy. Apollo Cradle and Children's Hospital. URL: https://www.apollocradle.com/blog/pregnancy/all-about-providing-emotional-support-to-women-during-pregnancy [accessed 2024-04-09]
3. Akesson B. Addressing the psychosocial needs of pregnant women affected by war: program approaches and program gaps. Refuge 2008;25(1):55-59. [doi: 10.25071/1920-7336.21397]
4. Sarkar NDP, Bardaji A, Peeters Grietens K, Bunders-Aelen J, Baingana F, Criel B. The social nature of perceived illness representations of perinatal depression in rural Uganda. Int J Environ Res Public Health 2018 Jun 7;15(6):1197. [doi: 10.3390/ijerph15061197] [Medline: 29880729]
5. Lund C, Town C. Maternal depression. : Health & Education Advice & Resource Team; 2016 Apr URL: https://d1wqtxts1xzle7.cloudfront.net/81141598/Crick-Lund-Reading-Pack-libre.pdf?1645431793=&response-content-disposition=inline%3B+filename%3DMaternal_Depression. [accessed 2024-12-20]
6. Tol WA, Ebrecht B, Aiyo R, et al. Maternal mental health priorities, help-seeking behaviors, and resources in post-conflict settings: a qualitative study in eastern Uganda. BMC Psychiatry 2018 Feb 7;18(1):39. [doi: 10.1186/s12888-018-1626-x] [Medline: 29415710]
7. Baron EC, Hanlon C, Mall S, et al. Maternal mental health in primary care in five low- and middle-income countries: a situational analysis. BMC Health Serv Res 2016 Feb 16;16:53. [doi: 10.1186/s12913-016-1291-z] [Medline: 26880075]

XSL•FO
RenderX

8. Muzik M, Borovska S. Perinatal depression: implications for child mental health. Ment Health Fam Med 2010 Dec;7(4):239-247. [Medline: 22477948]
9. Nwoke CN, Awosoga OA, McDonald S, Bonifacio GT, Leung BMY. African immigrant mothers' views of perinatal mental health and acceptability of perinatal mental health screening: quantitative cross-sectional survey study. JMIR Form Res 2023 Jan 27;7:e40008. [doi: 10.2196/40008] [Medline: 36705944]
10. Fisher J, Cabral de Mello M, Patel V, et al. Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review. Bull World Health Organ 2012 Feb 1;90(2):139-149H. [doi: 10.2471/BLT.11.091850]
11. Nemoda Z, Szyf M. Epigenetic alterations and prenatal maternal depression. Birth Defects Res 2017 Jul 17;109(12):888-897. [doi: 10.1002/bdr2.1081] [Medline: 28714605]
12. Field J. Good for your soul? Adult learning and mental well‐being. Int J Lifelong Educ 2009 Mar;28(2):175-191. [doi: 10.1080/02601370902757034]
13. de Bruijn A, van Bakel HJA, van Baar AL. Sex differences in the relation between prenatal maternal emotional complaints and child outcome. Early Hum Dev 2009 May;85(5):319-324. [doi: 10.1016/j.earlhumdev.2008.12.009] [Medline: 19162414]
14. Atukunda P, Muhoozi GKM, Westerberg AC, Iversen PO. Nutrition, hygiene and stimulation education for impoverished mothers in rural Uganda: effect on maternal depression symptoms and their associations to child development outcomes. Nutrients 2019 Jul 11;11(7):1561. [doi: 10.3390/nu11071561] [Medline: 31373314]
15. Krohn H, Guintivano J, Frische R, Steed J, Rackers H, Meltzer-Brody S. App-based ecological momentary assessment to enhance clinical care for postpartum depression: pilot acceptability study. JMIR Form Res 2022 Mar 23;6(3):e28081. [doi: 10.2196/28081] [Medline: 35319483]
16. Shi P, Ren H, Li H, Dai Q. Maternal depression and suicide at immediate prenatal and early postpartum periods and psychosocial risk factors. Psychiatry Res 2018 Mar;261:298-306. [doi: 10.1016/j.psychres.2017.12.085] [Medline: 29331710]
17. Patel V, Saxena S, Lund C, et al. The Lancet Commission on global mental health and sustainable development. Lancet 2018 Oct;392(10157):1553-1598. [doi: 10.1016/S0140-6736(18)31612-X] [Medline: 30314863]
18. Cox EQ, Sowa NA, Meltzer-Brody SE, Gaynes BN. The perinatal depression treatment cascade. J Clin Psychiatry 2016 Sep 28;77(09):1189-1200. [doi: 10.4088/JCP.15r10174]
19. Nakku JEM, Nakasi G, Mirembe F. Postpartum major depression at six weeks in primary health care: prevalence and associated factors. Afr Health Sci 2006 Dec;6(4):207-214. [doi: 10.5555/afhs.2006.6.4.207] [Medline: 17604509]
20. Ayano G, Demelash S, Yohannes Z, et al. Misdiagnosis, detection rate, and associated factors of severe psychiatric disorders in specialized psychiatry centers in Ethiopia. Ann Gen Psychiatry 2021 Feb 2;20(1):10. [doi: 10.1186/s12991-021-00333-7] [Medline: 33531016]
21. Altamura AC, Goikolea JM. Differential diagnoses and management strategies in patients with schizophrenia and bipolar disorder. Neuropsychiatr Dis Treat 2008 Feb;4(1):311-317. [doi: 10.2147/ndt.s2703] [Medline: 18728801]
22. Nasrallah HA. Consequences of misdiagnosis: inaccurate treatment and poor patient outcomes in bipolar disorder. J Clin Psychiatry 2015 Oct;76(10):e1328. [doi: 10.4088/JCP.14016tx2c] [Medline: 26528666]
23. Nakku JEM, Okello ES, Kizza D, et al. Perinatal mental health care in a rural African district, Uganda: a qualitative study of barriers, facilitators and needs. BMC Health Serv Res 2016 Jul 22;16(1):295. [doi: 10.1186/s12913-016-1547-7] [Medline: 27443346]
24. Muhwezi WW, Agren H, Musisi S. Detection of major depression in Ugandan primary health care settings using simple questions from a subjective well-being (SWB) subscale. Soc Psychiatry Psychiatr Epidemiol 2007 Jan;42(1):61-69. [doi: 10.1007/s00127-006-0132-5] [Medline: 17082896]
25. Rush JA. Handbook of Psychiatric Measures: American Psychiatric Association; 2000:166-168.
26. Spies G, Kader K, Kidd M, et al. Validity of the K-10 in detecting DSM-IV-defined depression and anxiety disorders among HIV-infected individuals. AIDS Care 2009 Sep;21(9):1163-1168. [doi: 10.1080/09540120902729965] [Medline: 20024776]
27. Baggaley RF, Ganaba R, Filippi V, et al. Detecting depression after pregnancy: the validity of the K10 and K6 in Burkina Faso. Trop Med Int Health 2007 Oct;12(10):1225-1229. [doi: 10.1111/j.1365-3156.2007.01906.x] [Medline: 17956505]
28. Kaaya SF, Fawzi MCS, Mbwambo JK, Lee B, Msamanga GI, Fawzi W. Validity of the Hopkins Symptom Checklist-25 amongst HIV-positive pregnant women in Tanzania. Acta Psychiatr Scand 2002 Jul;106(1):9-19. [doi: 10.1034/j.1600-0447.2002.01205.x] [Medline: 12100343]
29. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav Sci 1974 Jan;19(1):1-15. [doi: 10.1002/bs.3830190102] [Medline: 4808738]
30. Kushniruk AW, Myers K, Borycki EM, Kannry J. Exploring the relationship between training and usability: a study of the impact of usability testing on improving training and system deployment. Stud Health Technol Inform 2009;143:277-283. [doi: 10.3233/978-1-58603-979-0-277] [Medline: 19380948]
31. Chibanda D, Mangezi W, Tshimanga M, et al. Validation of the Edinburgh Postnatal Depression Scale among women in a high HIV prevalence area in urban Zimbabwe. Arch Womens Ment Health 2010 Jun;13(3):201-206. [doi: 10.1007/s00737-009-0073-6] [Medline: 19760051]

XSL•FO
RenderX

32. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The MINI-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20(20):22-33. [Medline: 9881538]
33. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150(6):782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
34. National health facility master list. Ministry of Health Republic of Uganda. 2018. URL: https://library.health.go.ug/health-infrastructure/health-facility-inventory/national-health-facility-master-facility-list-2018 [accessed 2024-12-27]
35. Matthey S. Using the Edinburgh Postnatal Depression Scale to screen for anxiety disorders. Depress Anxiety 2008;25(11):926-931. [doi: 10.1002/da.20415] [Medline: 18041072]
36. Asiki G, Baisley K, Newton R, et al. Adverse pregnancy outcomes in rural Uganda (1996-2013): trends and associated factors from serial cross sectional surveys. BMC Pregnancy Childbirth 2015 Oct 29;15(1):279. [doi: 10.1186/s12884-015-0708-8] [Medline: 26515763]
37. Moyer CA, Kolars CK, Oppong SA, Bakari A, Bell A, Busingye P. Predictors of stillbirths and neonatal deaths in rural western Uganda. Intl J Gynecology Obste 2016 Aug;134(2):190-193. [doi: 10.1016/j.ijgo.2016.01.009]
38. Kim HY. Statistical notes for clinical researchers: chi-squared test and Fisher's exact test. Restor Dent Endod 2017 May;42(2):152-155. [doi: 10.5395/rde.2017.42.2.152] [Medline: 28503482]
39. Rahman A, Iqbal Z, Harrington R. Life events, social support and depression in childbirth: perspectives from a rural community in the developing world. Psychol Med 2003 Nov;33(7):1161-1167. [doi: 10.1017/S0033291703008286]
40. Umuziga MP, Gishoma D, Hynie M, Nyirazinyoye L. Antenatal depressive symptoms in Rwanda: rates, risk factors, and social support. BMC Pregnancy Childbirth 2022 Mar 8;22(1):193. [doi: 10.1186/s12884-022-04522-4] [Medline: 35260093]
41. Yin X, Sun N, Jiang N, et al. Prevalence and associated factors of antenatal depression: systematic reviews and meta-analyses. Clin Psychol Rev 2021 Feb;83:101932. [doi: 10.1016/j.cpr.2020.101932] [Medline: 33176244]
42. Lee EJ, Lee JY, Lee SJ, Yu SE. Influence of self-esteem and spouse support on prenatal depression in pregnant women. J Korean Soc Matern Child Health 2020 Oct;24(4):212-220. [doi: 10.21896/jksmch.2020.24.4.212]
43. Hu Y, Wang Y, Wen S, et al. Association between social and family support and antenatal depression: a hospital-based study in Chengdu, China. BMC Pregnancy Childbirth 2019 Nov 19;19(1):420. [doi: 10.1186/s12884-019-2510-5] [Medline: 31744468]
44. Yang H, Zheng X, Zhou R, Shen Z, Huang X. Fertility behavior and depression among women: evidence from China. Front Psychol 2020;11:565508. [doi: 10.3389/fpsyg.2020.565508] [Medline: 33281665]
45. Kravdal Ø, Grundy E, Skirbekk V. Fertility history and use of antidepressant medication in late mid-life: a register-based analysis of Norwegian women and men. Aging Ment Health 2017 May;21(5):477-486. [doi: 10.1080/13607863.2015.1118010] [Medline: 26644174]
46. Golovina K, Elovainio M, Hakulinen C. Association between depression and the likelihood of having children: a nationwide register study in Finland. Am J Obstet Gynecol 2023 Feb;228(2):211. [doi: 10.1016/j.ajog.2022.10.016] [Medline: 36283480]
47. Li X, Gao R, Dai X, et al. The association between symptoms of depression during pregnancy and low birth weight: a prospective study. BMC Pregnancy Childbirth 2020 Mar 6;20(1):147. [doi: 10.1186/s12884-020-2842-1] [Medline: 32138708]
48. Muraca GM, Joseph KS. The association between maternal age and depression. J Obstet Gynaecol Can 2014 Sep;36(9):803-810. [doi: 10.1016/S1701-2163(15)30482-5] [Medline: 25222359]
49. Lee AM, Lam SK, Sze Mun Lau SM, Chong CSY, Chui HW, Fong DYT. Prevalence, course, and risk factors for antenatal anxiety and depression. Obstet Gynecol 2007 Nov;110(5):1102-1112. [doi: 10.1097/01.AOG.0000287065.59491.70] [Medline: 17978126]

Abbreviations

EPDS: Edinburgh Postnatal Depression Scale
KCCA: Kampala Capital City Authority

Edited by E Mensah; submitted 11.08.23; peer-reviewed by FN Abd Rahman, S Sawesi; revised version received 23.08.24; accepted 14.10.24; published 14.01.25.

Please cite as:
Namatovu HK, Magumba MA, Akena D
E-Screening for Prenatal Depression in Kampala, Uganda Using the Edinburgh Postnatal Depression Scale: Survey Results
Online J Public Health Inform 2025;17:e51602
URL: https://ojphi.jmir.org/2025/1/e51602
doi:10.2196/51602

XSL•FO
RenderX

© Hasifah Kasujja Namatovu, Mark Abraham Magumba, Dickens Akena. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 14.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Real-World Data on Alcohol Consumption Behavior Among Smartphone Health Care App Users in Japan: Retrospective Study

Kana Eguchi[1], PhD; Takeaki Kubota[2], MS; Tomoyoshi Koyanagi[3], PhD; Manabu Muto[4], MD, PhD

[1]Department of Informatics, Graduate School of Informatics, Kyoto University, Kyoto, Japan

[2]Department of Real World Data Research and Development, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[3]Business Development Office, Institute for Advancement of Clinical and Translational Science (iACT), Kyoto University Hospital, Kyoto, Japan

[4]Department of Therapeutic Oncology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Corresponding Author:
Kana Eguchi, PhD
Department of Informatics
Graduate School of Informatics
Kyoto University
Yoshida-Hommachi
Sakyo-ku,
Kyoto, 606-8501
Japan
Phone: 81 75 753 3369
Email: kana.eguchi@ieee.org

Abstract

Background: Although many studies have used smartphone apps to examine alcohol consumption, none have clearly delineated long-term (>1 year) consumption among the general population.

Objective: The objective of our study is to elucidate in detail the alcohol consumption behavior of alcohol drinkers in Japan using individual real-world data. During the state of emergency associated with the COVID-19 outbreak, the government requested that people restrict social gatherings and stay at home, so we hypothesize that alcohol consumption among Japanese working people decreased during this period due to the decrease in occasions for alcohol consumption. This analysis was only possible with individual real-world data. We also aimed to clarify the effects of digital interventions based on notifications about daily alcohol consumption.

Methods: We conducted a retrospective study targeting 5-year log data from January 1, 2018, to December 31, 2022, obtained from a commercial smartphone health care app (CALO mama Plus). First, to investigate the possible size of the real-world data, we investigated the rate of active users of this commercial smartphone app. Second, to validate the individual real-world data recorded in the app, we compared individual real-world data from 9991 randomly selected users with government-provided open data on the number of daily confirmed COVID-19 cases in Japan and with nationwide alcohol consumption data. To clarify the effects of digital interventions, we investigated the relationship between 2 types of notification records (ie, “good” and “bad”) and a 3-day daily alcohol consumption log following the notification. The protocol of this retrospective study was approved by the Ethics Committee of the Kyoto University Graduate School and Faculty of Medicine (R4699).

(Online J Public Health Inform 2025;17:e57084) doi:10.2196/57084

KEYWORDS

alcohol consumption; individual behavior; mobile health; mobile health app; mobile health care app log-based survey; real-world data; RWD; RWD analysis; smartphone health care app; surveillance system; health care app

Introduction

Alcohol consumption contributes to 3 million deaths globally each year, with harmful use responsible for 4.7% of the global burden of disease [1]. The World Health Organization recommends reducing harmful alcohol use by developing surveillance systems for alcohol consumption [2]. Although several studies using smartphone apps have been conducted [3], none have examined the general population’s long-term alcohol consumption behavior, that is, for a period of time longer than a year, nor the effect of digital interventions.

XSL•FO
RenderX

To elucidate detailed alcohol consumption behavior among alcohol drinkers in Japan, we conducted a retrospective study targeting 5-year logs obtained from a commercial smartphone health care app (CALO mama Plus [4]; Wellmira Inc). Common drinking occasions for Japanese working people include welcome events, get-togethers, farewell parties, and regular after-work drinking parties, so we hypothesized that there was a decrease in alcohol consumption during the period of the COVID-19 outbreak and the state of emergency, when individuals were requested to stay home and reduce social gatherings. We tested this hypothesis with individual real-world data; this analysis was only possible with such data. We aimed to clarify the effects of digital interventions using notifications and daily records of alcohol consumption.

Methods

Overview

We analyzed the smartphone app logs to investigate two perspectives: (1) user numbers and (2) alcohol consumption behavior (ie, mean net alcohol consumption and alcohol-related notifications, together with their effectiveness). Because our target period coincided with the COVID-19 outbreak, we compared these values with the daily number of confirmed COVID-19 cases in Japan [5].

Target Commercial Smartphone Health Care App

CALO mama Plus [4] is a mobile health care app for managing nutrition. It records various health-related data, including alcohol consumption, and has been used in several health-related studies based on real-world data [6-9]. Details on the app can be found in Multimedia Appendix 1.

Ethical Considerations

The dataset used in this research was collected by a service provider company for its own purposes. Legal consent for primary and secondary uses was obtained upon initial activation of the app. This retrospective study obtained no additional consent. All data were deidentified by the operating company (Wellmira Inc). The study protocol was approved by the Ethics Committee of the Kyoto University Graduate School and Faculty of Medicine (R4699).

Materials

Analysis 1 included all registered users of the target app. We calculated the rate of active users as those who input any data into the app as a proportion of the number of registered users. In other words, active users hypothetically included people who did not drink on that day and those who do not habitually drink alcohol (including nondrinkers). The rate of users inputting alcohol data was calculated as those who input alcohol consumption data to the application as a proportion of the number of registered users.

For analysis 2, we used alcohol consumption records from 9991 randomly selected users. The method for selecting target users is described in the Data Exclusion subsection. Table 1 describes the background information of the users. To confirm the validity of the data, we compared individual alcohol consumption to the nationwide alcohol consumption records published by the National Tax Agency of Japan [10].

Table 1. Background information for 9991 randomly selected users of the target commercial smartphone app (CALO mama Plus [4]). We randomly selected these users from those who logged alcohol consumption records more than 20 times in a year for at least one calendar year from January 1, 2018, to December 31, 2022, and we confirmed that their age and alcohol input were logically reasonable in light of Japanese law and human physiology. Age represents age at the time of data extraction on December 31, 2022; weight, height, and BMI represent values at the time a user registered on the app. *P* values were calculated with the Brunner-Munzel test.

	Male	Female	*P* value	W statistic [11]
Participants, n (%)	6612 (66.2%)	3379 (33.8%)	—[a]	—[a]
Age (years), mean (SD)	52.9 (9.58)	50.1 (10.5)	<.001	12.8
Weight (kg), mean (SD)	73.0 (10.9)	57.0 (9.11)	<.001	–104
Height (m), mean (SD)	1.72 (0.0586)	1.59 (0.0519)	<.001	–214
BMI (kg/m^2), mean (SD)	24.8 (3.32)	22.6 (3.36)	<.001	–60.9

[a]Not applicable.

Analytical Conditions in Analysis 2

Data Exclusion

Initially, we randomly extracted 10,000 users who logged alcohol consumption data more than 20 times in a year for at least one calendar year from January 1, 2018, to December 31, 2022.

Considering the legal drinking age in Japan (20 years) and the average life expectancy in Japan, we excluded 7 users whose ages were 2, 3, 4, 6, 9, 18, and 123 at the time of data extraction (ie, December 31, 2022). In addition, we excluded 2 users who recorded unrealistic alcohol amounts (eg, over 1000 g net alcohol/day). Thus, the final number of target users was 9991.

Notification Evaluation

In the effectiveness evaluation of the notifications in analysis 2, the app logs for net alcohol consumption and alcohol-related notifications were analyzed to investigate whether notifications influenced drinking behavior following the notification.

We focused on 2 alcohol-related notifications derived from net alcohol intake: “good” (<20 g/day) and “bad” (≥60 g/day). We considered a notification effective when either of two conditions was satisfied: (1) at least one “no drinking day” occurred within 3 days, or (2) there were no days with ≥60 g/day within 3 days.

XSL•FO
RenderX

The effectiveness rate was calculated by dividing effective notifications by the total number of notifications and multiplying by 100.

Results

Analysis 1

Figure 1 shows the trend of the number of app users. Although the exact number is undisclosed due to company restrictions, registered users increased 20-fold during the target period. The rate of active users and users with alcohol input data remained stable over this period, with average rates of 5.6% (SD 3.19%) and 1.17% (SD 0.689%), respectively. Compared with the number of daily confirmed COVID-19 cases in Japan (Figure 2), no marked changes were confirmed in any of these values.

Figure 1. Trends among users of the target commercial smartphone app (CALO mama Plus [4]) with regard to the number of registered users, rate of active users, and rate of users with alcohol input data for the target period (ie, from January 1, 2018, to December 31, 2022). The solid black line indicates the date of the first state of emergency in Japan, starting on April 7, 2020, in key prefectures (Tokyo, Kanagawa, Saitama, Chiba, Osaka, Hyogo, and Fukuoka), which then expanded nationwide from April 16, 2020 [12]. The gray lines indicate the date of media exposure on TV programs. Although the exact number of users is undisclosed due to company restrictions, registered users increased 20-fold during the target 5 years. However, the rate of active users and users with alcohol input data remained stable over this period, with average rates of 5.6% (SD 3.19%) and 1.17% (SD 0.689%), respectively.

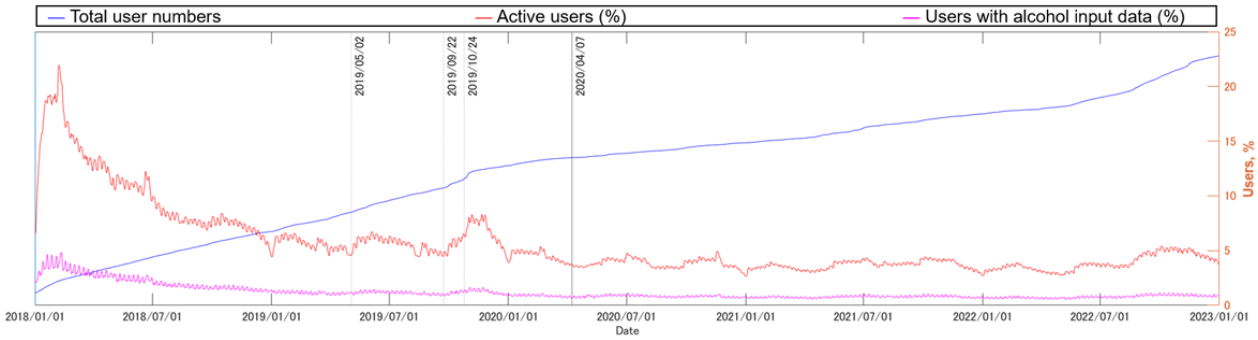
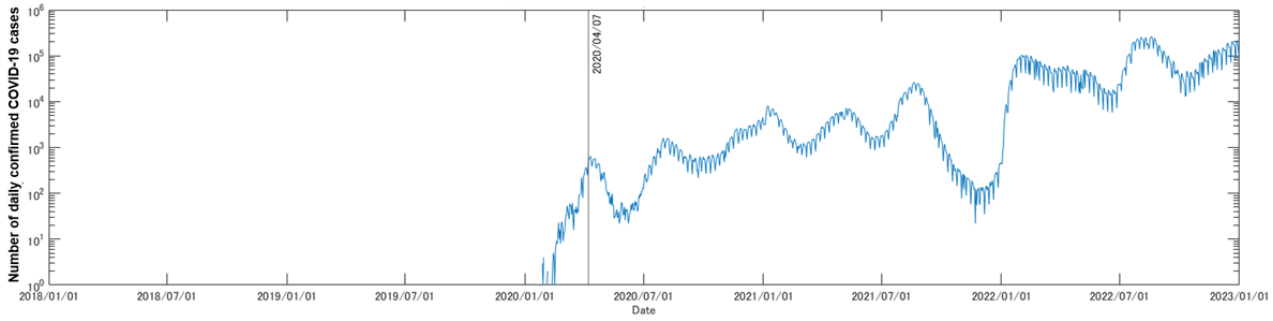


Figure 2. The number of daily confirmed COVID-19 cases in Japan [5] for the target period (ie, from January 1, 2018, to December 31, 2022). For easier understanding, we used logarithmic values in the y-axis. The solid black line indicates the date of the first state of emergency in Japan, starting on April 7, 2020, in key prefectures (Tokyo, Kanagawa, Saitama, Chiba, Osaka, Hyogo, and Fukuoka), which then expanded nationwide from April 16, 2020 [12]. We can confirm that this date was approximately the peak of the first wave of the COVID-19 outbreak in Japan.



Analysis 2

Figure 3 shows Japan's alcohol consumption trends, published by the National Tax Agency of Japan [10]. Overall, total alcohol consumption declined, with an annual drop of about 100,000 kL after 2018. Notably, there was a decrease of 300,000 kL between 2019 and 2020, and the total alcohol consumption was similar in fiscal years 2020, 2021, and 2022. Given that fiscal years 2020 and 2021 included the state of emergency due to the COVID-19 outbreak [12], this gap may reflect the government's request to limit social gatherings.

XSL•FO
RenderX

Figure 3. Alcohol consumption trends from the nationwide record published by the National Tax Agency of Japan [10]. These records are tallied according to the Japanese fiscal year (April 1 to March 31). "Mirin" (sweet sake) is not an alcoholic beverage, but a traditional Japanese sweet cooking wine containing approximately 10% alcohol. Overall, total alcohol consumption declined, with an annual drop of about 100,000 kL after 2018. Notably, there was a decrease of 300,000 kL between 2019 and 2020, and the amount of total alcohol consumption was similar in fiscal years 2020, 2021, and 2022.

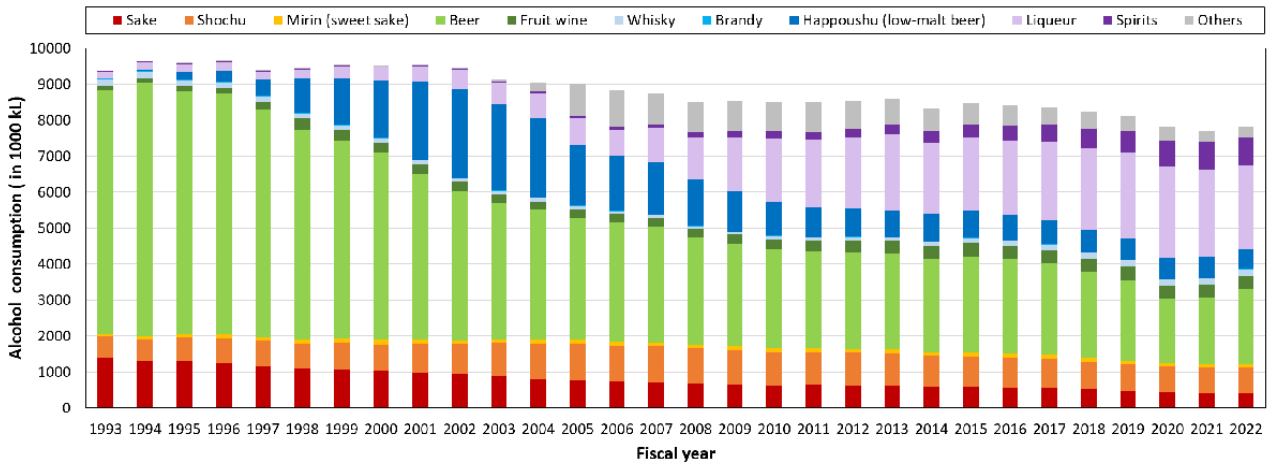
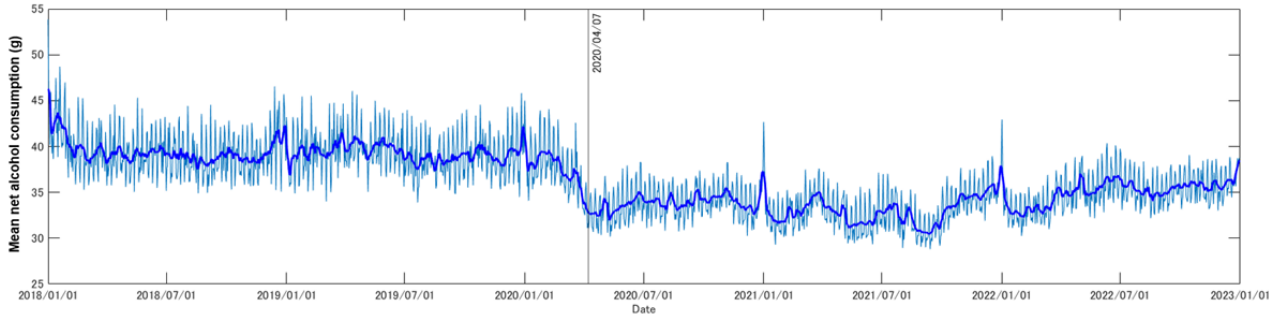


Figure 4 shows the trend in mean net alcohol consumption. Figure 5 shows the number and effectiveness of "good" notifications, whereas Figure 6 shows that of "bad" notifications. Notably, both mean net alcohol consumption (Figure 4) and the effect of "bad" alcohol-related notifications (Figure 6) decreased at the beginning of April 2020. Based on the Brunner-Munzel test, in the periods from January 1, 2018, to March 31, 2020, and April 1, 2020, to December 31, 2022, these differences were confirmed as significant (net alcohol consumption: W statistic=90.8; P<.001; effect of "bad" alcohol-related notifications: W statistic=34.4; P<.001). The Brunner-Munzel test results for other factors are shown in Figures 4-6, and other details can be found in Multimedia Appendix 1.

Figure 4. Mean net alcohol consumption obtained from the logs of the target commercial smartphone app (CALO mama Plus [4]) for the target period (January 1, 2018, to December 31, 2022). The bold blue line indicates the 7-day moving average. The solid black line indicates the date of the first state of emergency in Japan, starting on April 7, 2020, in key prefectures (Tokyo, Kanagawa, Saitama, Chiba, Osaka, Hyogo, and Fukuoka), which then expanded nationwide from April 16, 2020 [12]. Around the date of the first state of emergency in Japan, the data confirm a rapid decline. Aside from this, in all target years, the data confirm temporal peaks around the year end and the New Year holiday, which can be related to the increase in drinking parties at such times.



XSL•FO
RenderX

Figure 5. Number of "good" notifications, together with their effectiveness, obtained from the logs of the target commercial smartphone app (CALO mama Plus [4]) for the target period (January 1, 2018, to December 31, 2022). The pale blue line indicates the number of "good" notifications, whereas the pale red line indicates the effectiveness rate. The bold lines indicate the 7-day moving averages. The solid black line indicates the date of the first state of emergency in Japan, starting on April 7, 2020, in key prefectures (Tokyo, Kanagawa, Saitama, Chiba, Osaka, Hyogo, and Fukuoka), which then expanded nationwide from April 16, 2020 [12]. Unlike "bad" notifications, there were no marked changes around the date of the first state of emergency in Japan in either the number of notifications or their effectiveness.

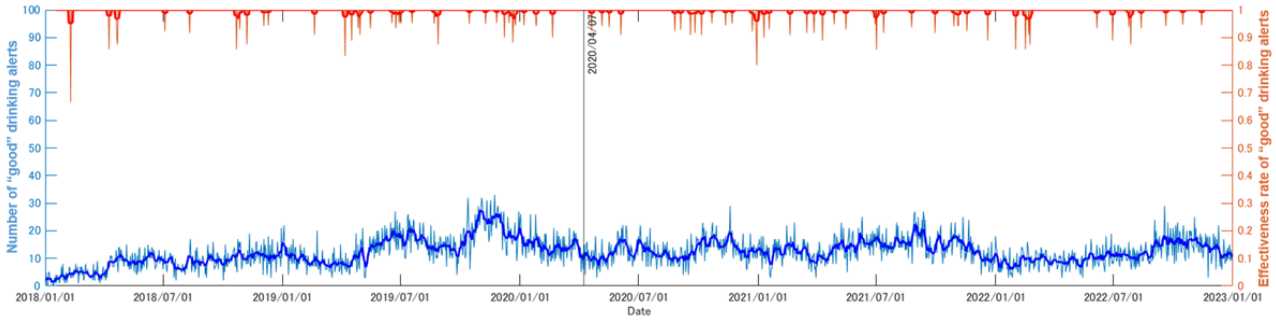
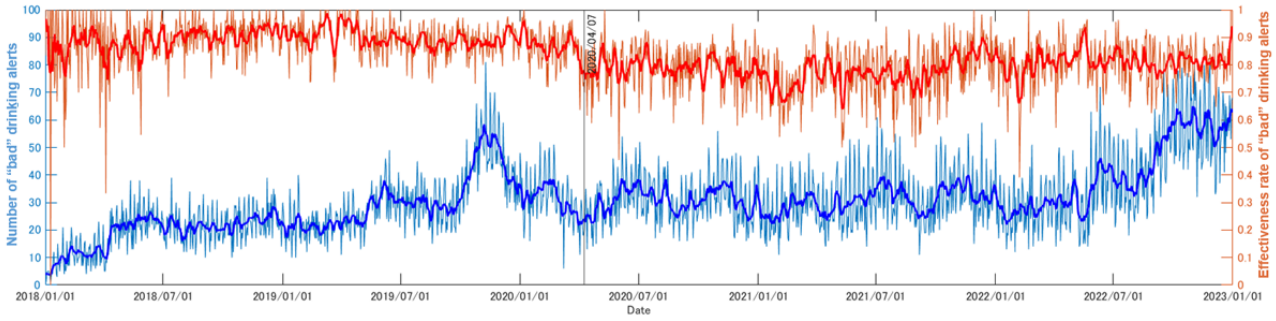


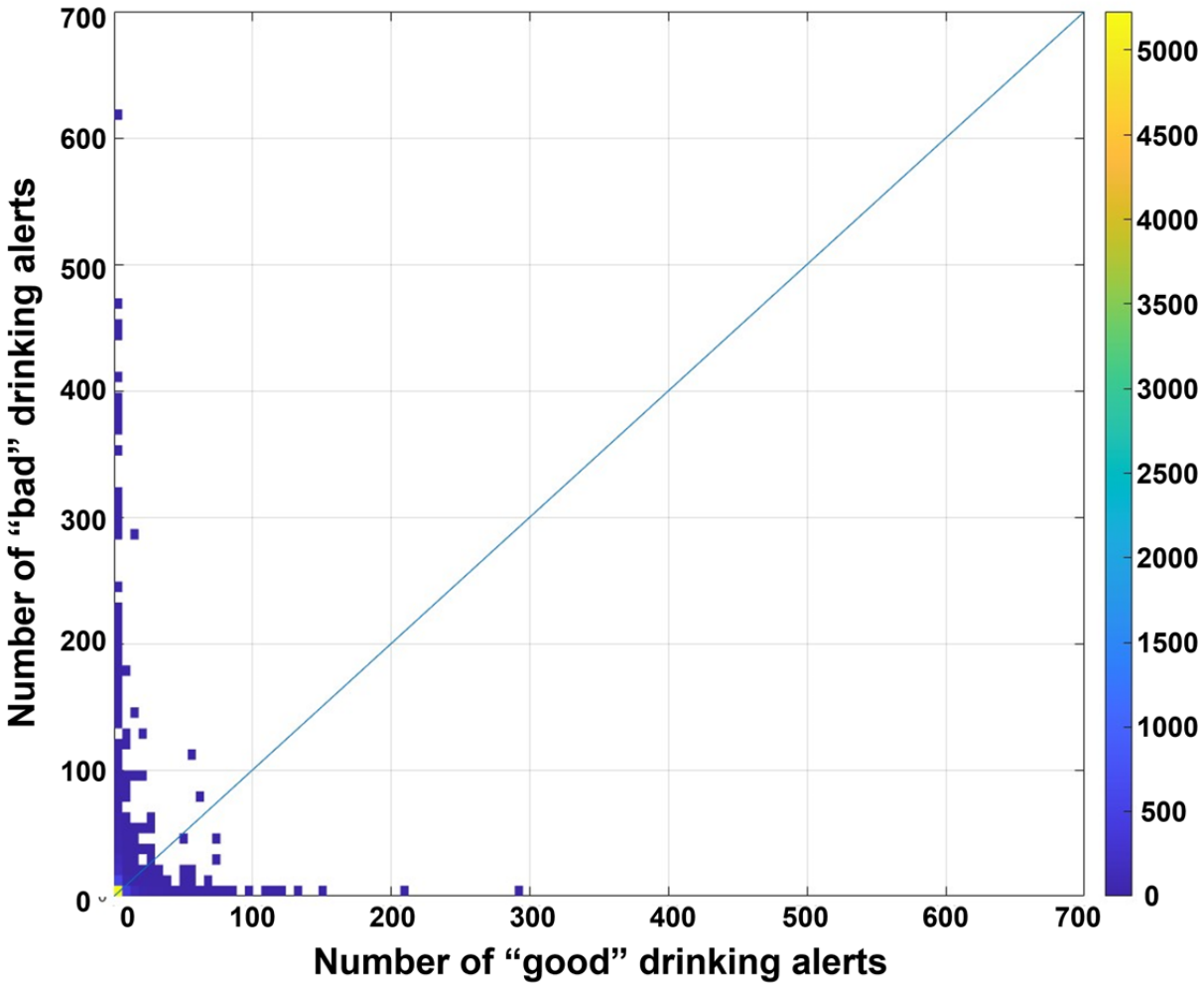
Figure 6. Number of "bad" notifications and their effectiveness, obtained from the logs of the target commercial smartphone app (CALO mama Plus [4]) for the target period (ie, from January 1, 2018, to December 31, 2022). The pale blue line indicates the number of "bad" notifications, whereas the pale red line indicates their effectiveness rate. The bold lines indicates the 7-day moving averages. The solid black line indicates the date of the first state of emergency in Japan starting on April 7, 2020, in key prefectures (Tokyo, Kanagawa, Saitama, Chiba, Osaka, Hyogo, and Fukuoka), which then expanded nationwide from April 16, 2020 [12]. We can confirm that there was a rapid decline in the effectiveness of the notifications near the date of the first state of emergency in Japan, by approximately 10%.



Comparing "good" (Figure 5) and "bad" (Figure 6) notifications, the number of "bad" notifications was roughly double. A 2D histogram (Figure 7) depicts the balance between "good" (x-axis) and "bad" notifications (y-axis) for each individual. This also confirms that "bad" notifications were more common. Accordingly, target users (n=9991) were classified into four groups: (1) more "bad" than "good" (n=4180, 41.8%); (2) equal (n=369, 3.69%); (3) more "good" than "bad" (n=2787, 27.9%); and (4) no notifications (n=2655, 26.6%). In groups 1, 2, and 3, a total of 6122 (61.3%) users had <10 "bad" notifications; thus, 8777 (87.8%) of all target users had no or <10 "bad" notifications. Among the remaining 1214 (12.2%) users, 1176 (96.9%) were classified as being in group 1 (more "bad" than "good"). Thus, only specific users, such as heavy drinkers, possibly tended to receive "bad" notifications.

XSL•FO
RenderX

Figure 7. 2D histogram of "good" drinking notifications and "bad" drinking notifications obtained from the logs of the target commercial smartphone app (CALO mama Plus [4]) for the target period (January 1, 2018, to December 31, 2022). The blue line indicates x = y, namely, that there were equal numbers of "good" and "bad" drinking notifications. The dots in the graph below the x = y line indicate the people who received more "good" notifications than "bad" and vice versa. Regarding the target users in our study, more people tended to have more "bad" notifications than "good" ones (ie, the dots in the area of the graph above the x = y line).



Discussion

Summary of Findings

We confirmed that there was a decrease in net alcohol consumption in both government-provided data (Figure 3) and individual real-world data (Figure 4) that continued after 2020, which may have been related to the outbreak of COVID-19 and the state of emergency in Japan. Regarding the effect of the digital interventions, only the "bad" notifications showed a decreased effect near the first state of emergency. Taken together, this suggests that people who only consume alcohol at social gatherings may have reduced their consumption, whereas those who originally had no interest in "bad" notifications continued their consumption.

Compared with government-provided data, the individual real-world data showed more details on alcohol consumption behavior. In all target years, we confirmed temporal peaks in mean net alcohol consumption around the year end and the New Year holiday, which might be related to the increase in drinking parties at these times. This may paradoxically support our hypothesis that the request for restricting social gatherings led to reduced alcohol consumption in Japan.

Limitations

The main limitations of this study are its generalizability and reliability. As we showed in Figure 1, the 5-year rates of active app users and users who input alcohol data were both lower than 10%. This low active user rate is one of the peculiarities of real-world data, indicating that our results represent only one aspect of reality.

Another limitation is the app specifications (detailed in Multimedia Appendix 1). Comparing alcohol consumption and notification records, we found "good" notifications in only 4.61% of records meeting the criterion of <20 g net alcohol/day, whereas "bad" notifications were found in 27.5% of records meeting the criterion of ≥60 g net alcohol/day. Therefore, our notification analysis can be considered a pilot study, and the

XSL•FO
RenderX

efficacy of notifications should be assessed independently in a prospective study with modified alert systems.

The final limitation is the influence of the COVID-19 outbreak. Because almost half of our target period included the COVID-19 outbreak, our analyses may be affected by various socioeconomic factors. Further investigation focusing on individual behavior change is required to elucidate detailed relationships.

Conclusions

A 5-year analysis of individual real-world data, including data from the COVID-19 outbreak period, showed that real-world data may reflect a more detailed reality than government-provided data.

Acknowledgments

This research was supported by JST START University Ecosystem Promotion Type (Supporting Creation of Startup Ecosystem in Startup Cities; grant JPMJST2181), Japan, and by the startup business support program IPG-Advance from the Office of Society-Academia Collaboration for Innovation (SACI), Kyoto University. This work was also supported by the research fund of the Department of Real World Data Research and Development, Graduate School of Informatics, Kyoto University, which is funded by joint research contributions from Nippon Telegraph and Telephone (NTT) Corporation, Canon Medical Systems Corporation, H.U. Group Holdings, Inc, and Ark Medical Solutions Inc to Kyoto University. The authors thank Keiko Ishimura, Shinichiro Mizuno, Ryusuke Miki, Ayako Saijo, Xie Chen, and Yuki Sasaki at Wellmira Inc for their technical support.

Data Availability

The real-world datasets on alcohol consumption analyzed during this study are not available as they are confidential business data of Wellmira Inc, but they are available from the corresponding author on reasonable request. The datasets for the number of daily confirmed COVID-19 cases in Japan are available from the Ministry of Health, Labor and Welfare in Japan [5], whereas the datasets for alcohol consumption trends are available from the National Tax Agency of Japan [10].

Authors' Contributions

Conceptualization: KE, T Kubota, T Koyanagi, MM
Data curation: KE, T Kubota
Formal analysis: KE
Project administration: T Koyanagi, MM
Software: KE
Visualization: KE
Writing—original draft preparation: KE
Writing—review and editing: KE, T Kubota, T Koyanagi, MM

Conflicts of Interest

KE belonged to the Department of Real World Data Research and Development, Graduate School of Informatics, Kyoto University, from July 2022 to March 2024. T Koyanagi has served as an advisor for Nippon Telegraph and Telephone Corporation (NTT), Academic Industry Research Inc, and Kyoto University Innovation Capital Co, Ltd. Additionally, he holds leadership and advisory positions, including serving as the representative director of the General Incorporated Association LSP, chairperson of the Non-Profit Organization Kyoto Commons, member of the Selection Committee for the Japan Medical Research and Development Grand Prize, member of the Project Evaluation Committee at the Japan Agency for Medical Research and Development (AMED), member of the Researcher Support and Grant Selection Committee at the Japan Pharmaceutical Manufacturers Association (JPMA), and member of the Scholarship Selection Committee at the Shuto Scholarship Foundation. These affiliations have not influenced the content or conclusions of this manuscript.

Multimedia Appendix 1
Supplementary remarks and analysis.
[DOCX File , 22 KB - ojphi_v17i1e57084_app1.docx]

References

1. Alcohol. World Health Organization. URL: https://www.who.int/health-topics/alcohol [accessed 2025-03-04]
2. Infographic: alcohol and health. World Health Organization. URL: https://www.who.int/images/default-source/departments/substances-abuse/alcohol/infographics/alcohol-3-million-death-every-year.png?sfvrsn=8062967_2 [accessed 2025-03-04]
3. Colbert S, Thornton L, Richmond R. Smartphone apps for managing alcohol consumption: a literature review. Addict Sci Clin Pract 2020 May 07;15(1):17 [FREE Full text] [doi: 10.1186/s13722-020-00190-x] [Medline: 32381062]
4. CALO Mama Plus. Wellmira Inc. URL: https://calomama.com/ [accessed 2025-03-04]

XSL•FO
RenderX

5. Open data on COVID-19. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/covid-19/open-data.html [accessed 2025-03-04]
6. Sato K, Sakata R, Murayama C, Yamaguchi M, Matsuoka Y, Kondo N. Changes in work and life patterns associated with depressive symptoms during the COVID-19 pandemic: an observational study of health app (CALO Mama) users. Occup Environ Med 2021 Sep;78(9):632-637 [FREE Full text] [doi: 10.1136/oemed-2020-106945] [Medline: 33619124]
7. Sasaki Y, Sato K, Kobayashi S, Asakura K. Nutrient and food group prediction as orchestrated by an automated image recognition system in a smartphone app (CALO mama): validation study. JMIR Form Res 2022 Jan 10;6(1):e31875 [FREE Full text] [doi: 10.2196/31875] [Medline: 35006077]
8. Nakata Y, Sasai H, Gosho M, Kobayashi H, Shi Y, Ohigashi T, et al. A smartphone healthcare application, CALO mama Plus, to promote weight loss: a randomized controlled trial. Nutrients 2022 Nov 02;14(21):4608 [FREE Full text] [doi: 10.3390/nu14214608] [Medline: 36364870]
9. Eguchi A, Kawamura Y, Kawashima T, Ghaznavi C, Ishimura K, Kohsaka S, et al. The efficacy of an mHealth app in facilitating weight loss among Japanese fitness center members: regression analysis study. JMIR Form Res 2023 Nov 08;7:e48435 [FREE Full text] [doi: 10.2196/48435] [Medline: 37938885]
10. Alcohol guidebook (June 2024). National Tax Agency Japan. URL: https://www.nta.go.jp/taxes/sake/shiori-gaikyo/shiori/2024/index.htm [accessed 2025-03-04]
11. scipy.stats. brunnermunzel. SciPy. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.brunnermunzel.html [accessed 2025-03-10]
12. Transition of the state of emergency for COVID-19. Digital Agency Japan. URL: https://hakusyo1.moj.go.jp/jp/69/nfm/n69_2_7_2_0_3.html [accessed 2025-03-04]

Edited by E Mensah; submitted 05.02.24; peer-reviewed by T Takemura, K Aguiar; comments to author 05.03.24; revised version received 13.05.24; accepted 26.02.25; published 25.03.25.

Please cite as:
Eguchi K, Kubota T, Koyanagi T, Muto M
Real-World Data on Alcohol Consumption Behavior Among Smartphone Health Care App Users in Japan: Retrospective Study
Online J Public Health Inform 2025;17:e57084
URL: https://ojphi.jmir.org/2025/1/e57084
doi:10.2196/57084
PMID:40131328

©Kana Eguchi, Takeaki Kubota, Tomoyoshi Koyanagi, Manabu Muto. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Accuracy of Nutrition-Related Awareness Messages on Twitter (Rebranded as X) by the Nutrition Awareness Providers in the Kingdom of Saudi Arabia: Validity Content Analysis

Duaa Alammari[1,2], MAS, DrPH; Iman A Bindayel[3], PhD; Noura Althukair[3], MSc; Noura AlRomi[3], BSc; Khalid Aldubayan[3], PhD; Najla Khateeb[4], PhD; Ahmed Alabdrabalnabi[5], DrPH; Banan Banamah, MPH

[1]Department of Health Systems Management, College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Prince Mutib Ibn Abdullah ibn Abdulaziz Rd, Riyadh, Saudi Arabia

[2]King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

[3]Department of Community Health and Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

[4]College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences, Al Ahsa, Saudi Arabia

[5]Department of Public Health, College of Health Sciences, Saudi Electronic University, Dammam, Saudi Arabia

Corresponding Author:
Duaa Alammari, MAS, DrPH
Department of Health Systems Management, College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Prince Mutib Ibn Abdullah ibn Abdulaziz Rd, Riyadh, Saudi Arabia

Related Article:

This is a corrected version. See correction statement: https://ojphi.jmir.org/2025/1/e85577

Abstract

Background: With the increasing use of social media, platforms like Twitter (rebranded as X in 2023) have become popular channels for disseminating health information. In Saudi Arabia, Twitter is widely used, making it an effective tool for health awareness. However, the accuracy of nutrition-related content on social media is often questioned.

Objective: The study aims to evaluate the accuracy and evidence-based quality of nutrition-related tweets posted by reputable Saudi health and nutrition awareness providers.

Methods: A mixed methods content analysis was conducted on tweets published by 7 Saudi health organizations, examining content in Arabic and English over 12 months. Nutrition-related tweets were analyzed for accuracy, popularity, and evidence inclusion by a panel of experts in clinical nutrition, food science, and technology.

Results: A total of 531 nutrition-related tweets were included in the study. Findings indicate that 445 (84%) of the tweets were accurate, of which only 17 (4%) included cited evidence. Yet, only 13 (2%) were inaccurate. The highest number of tweets are from Saudi Food and Drug Administration (SFDA) 96 (18%), Gulf Health Council (GHC) 91 (17%), Saudi Society for Clinical Nutrition (SSCN) 89 (16%), Kayl Association for Combating Obesity 83 (16%) and National Nutrition Committee (NNC) 80 (15%) and the lowest is Ministry of Health (MOH) 31 (5%). Significant relationships were observed between tweet accuracy and the source organization (P=.009, 95% CI 0.008 - 0.01), content type (P=.03, 95% CI 0.03 - 0.03), and tweet timing (P=.04, 95% CI 0.04 - 0.04). Governmental sources had higher popularity and were more frequently accurate compared to nongovernmental sources.

Conclusions: Reputable Twitter accounts in Saudi Arabia generally provide accurate nutrition-related content, though evidence citation is minimal. Users are encouraged to rely on reputable accounts for health information, and further research is suggested to explore the quality of evidence in such posts.

(Online J Public Health Inform 2025;17:e68128) doi:10.2196/68128

KEYWORDS

nutrition; awareness; content analysis; social media; twitter (X); nutritional; dietary; nutrition awareness; Asian; Saudi Arabian; media; health information; mixed methods; clinical nutrition; food science

XSL•FO
RenderX

Introduction

The global use of social media has increased exponentially, becoming one of the most popular online activities worldwide [1]. In 2020, over 3.6 billion people were using social media globally [2]. In line with this global trend, The Communications, Space, and Technology Commission (CST) has published the Saudi Internet Report 2022, indicating that there are 98.6% Internet use among the Saudi population [3]. The Ministry of Communications and Information Technology in Saudi Arabia reported a significant increase in the number of Saudi users of social programs and applications, rising from 8.5 million in 2015 to 18.3 million in 2020 [4]. Based on the Saudi Youth Statistic Report in 2023, there were 94.9 % of males and 94% of females participating on social networks [5]. Saudi Arabia was among the top 15 countries using Twitter (rebranded as X in July 2023), with over 12.45 million users [2,4], more than one-third of the Saudi population. Twitter ranks among the most popular social media platforms. It was launched in 2006 offering a free and interactive micro-blogging service where users create and share messages of up to 140 characters, called “tweets.” [6,7] There are over 500 million active users on Twitter alone, writing more than 340 million tweets and 1.6 billion search queries daily. Tweets can be posted via the web, instant online message, or mobile phone [6,7].

Social media is becoming a valuable tool for communication. Therefore, it is used in many fields, such as education, marketing, and health communication. The use of social media is limitless; some people use the platforms to receive medical consultations, while others use it to look for information related to their conditions [8]. Social media also provides a promising opportunity for public health promotion. For instance, social media platforms were used to deliver content as part of larger multiple component interventions for weight management [9]. It is considered a powerful tool for health education due to its ability to reach a large audience seamlessly, which makes it a practical and cost-effective tool to disseminate awareness content and messages [10,11]. Promoting awareness of dietary intake is vital for a healthy lifestyle and developing successful nutrition interventions. It is also considered the first step in population behavioral change [12,13]. Proper nutrition helps the body maintain healthy functions, growth, and reproduction and prevents diseases [12]. Participants mostly used Twitter to search for general knowledge, disease prevention and treatments for illness, nutrition, weight loss, and healthy lifestyle [8].

In recent years, patients have been engaging in unrestricted sharing of medical and health care information on social media [14]. Many patients use social media and online resources such as Facebook, Instagram, and Twitter to obtain health-related information to guide their health care decisions [14,15]. The widespread use and availability of social media in healthcare are attributed to several factors, such as the increase in ownership of mobile phones and the uptake of personal communication technologies worldwide [16]. The popularity of social media is perhaps due to its approachable and speedy nature and the ability to exchange user comments and feedback. In addition, governmental and health organizations have accounts on social media to disseminate news and health information to social media-using citizens.

According to the literature, the most popular platform studied for health-related information is Twitter [17]. Alassiri et al [8], mentioned that the number of Twitter users searching for health information in Saudi Arabia is increasing, two-thirds of the respondents preferred Twitter as a source due to its ease of use and accessibility. In Saudi Arabia, a study indicated that half of the participants considered Twitter their favorite social media platform, and a quarter of the participants indicated they preferred it to look for nutritional information [18].

Twitter is widespread and accepted in Saudi Arabia. The literature suggested that half of the Saudi Arabian population expressed a positive attitude toward seeking health information through Twitter [8]. Over a third of participants in a Saudi study expressed their satisfaction with using Twitter to find nutritional information [18]. Some participants reported having a better diet due to information available on Twitter [8] or that available health information on social media could enhance their health awareness [14].

Despite the numerous advantages that social media offers in the field of health communication, it has several challenging disadvantages. These challenges may include spreading falsified posts, misinformation, and the availability of massive amounts of information, which can create confusion among the general public [19,20]. Misinformed individuals can suffer from severe consequences that can lead to death in extreme cases as a result of social media rumors [20]. However, as the evidence suggests, the high number of tweets is not necessarily accurate [16,19]. A study indicated that tweets did not need to be accurate or evidence-based to gain traction [16]. While several studies have been conducted on the accuracy of health-related tweets, only 2 studies were conducted in Saudi Arabia by Albalawi et al [16] and Alnemer et al [6] focusing broadly on health, showing that at least half of the tweets are evidence-based. To our knowledge, no studies have specifically examined the content accuracy of nutrition-related tweets in Saudi Arabia, highlighting the gap that our study aims to address.

The study aims to evaluate the accuracy (evidence-based) of nutrition-related content on Twitter, posted by reputable Saudi Twitter accounts that are considered awareness providers disseminating nutritional awareness information in Saudi Arabia. Awareness providers are considered a reliable source of nutrition-related information, expecting their Twitter accounts and posts to hold the same level of reliability and accuracy. This study focuses on 7 Saudi Twitter accounts (governmental and nongovernmental) that provide and disseminate nutritional awareness content in the Kingdom of Saudi Arabia. Conducting such a study is a priority, considering the high popularity of Twitter in the Arab world.

Methods

This study examined tweets and retweets in Arabic or English with diet or nutritional content posted on the official Twitter accounts of nutritional awareness providers in Saudi Arabia

XSL•FO
RenderX

(governmental and nongovernmental). The study examined only official Twitter accounts affiliated with the entity to collect data.

This study performed a mixed methods content analysis of nutritional-related content published on Twitter. A semiqualitative approach was used to evaluate the content and accuracy of tweets. Twitter was chosen to investigate the nutrition-related content due to its popularity as a widely used social media platform in Saudi Arabia. This study followed a similar protocol to previous studies that assessed health-related content on Twitter [6,16]. A manual approach was used to identify and categorize nutritional-related tweets.

Identification of Tweets

Manual data collection took place on the social media platform Twitter, including tweets in Arabic or English related to diet and nutrition from 7 selected nutrition awareness providers in Saudi Arabia. The process began with a preliminary search to identify official accounts disseminating nutritional information. This search reviewed publicly recognized accounts of health authorities, educational institutions, and professional organizations known for sharing nutrition-related content. The selection focused on highly relevant accounts, activity levels, and credibility. Tweets related to diet and nutrition that met the predefined inclusion criteria were included in the extraction sheet (see Multimedia Appendix 1 for the full extraction sheet).

Criteria for Selecting Tweets (Inclusion Criteria)

The inclusion criteria for selecting tweets are as follows:

1. Any tweets focusing on nutrition-related content (text, media, graphics, or link to file).
2. Posted by 1 of the 7 nutrition awareness providers in Saudi Arabia that were identified by a preceding study. The study used 47 nutrition-related terms to search Twitter using the advanced search setting. The selection of the popular accounts was based on a minimum number of likes to 200, a minimum number of retweets to 200, and Saudi Arabia as a location. The following were identified as the most popular nutritional awareness providers in Saudi Arabia:

- Ministry of Health (MOH)
- Saudi Food and Drug Administration (SFDA)
- Gulf Health Council (GHC)
- Public Health Authority (Weqaya)
- Kayl Association for Combating Obesity
- Saudi Society for Clinical Nutrition (SSCN)
- National Nutrition Committee (NNC)
- Date of tweets starting from September 2020 until August 2021 (12-month period)

Data Collection

A total of 8 data collectors manually identified and screened nutrition-related tweets published on the 7 official Twitter accounts within 12 months (September 2020 to August 2021). The extraction of tweets occurred in 2 phases.

Phase 1

Included the manual identification of tweets by 8 reviewers. In addition, 2 members independently reviewed each official account to identify the 400 most recent nutrition-related tweets and retweets. First, 1 member identified and collected all nutritional tweets on the account by adding them to the extraction sheet. Second, another reviewer went through the Twitter account and added any tweets not previously included in the extraction sheet (first screening). If there were less than 400 tweets, all nutrition-related tweets are included for the 12 months. The results of identified tweets by the 2 reviewers were combined to ensure the inclusion of all nutrition-related tweets. Finally, a third reviewer screened the extracted list of tweets to ensure that all included tweets met the inclusion and exclusion criteria (second screening). This phase produced a total of 814 selected nutrition-related tweets. See Table 1 for details on nutritional tweets data collection and screening.

XSL•FO
RenderX

Table . Nutrition-related tweets data collected and screened per account.

Phase	Names of entities							Total tweets (N=531)
	NNC[a]	SSCN[b]	Kayl[c]	Weqaya[d]	GHC[e]	SFDA[f]	MOH[g]	
Tweets collected, n	71	172	101	68	206	173	23	814
1st screening	+13	+18	+29	+10	+22	+21	+19	+132
2nd screening	−1	−10	−40	−13	−27	−31	−11	−133
Total of overall included tweets, n	83	180	90	65	201	163	31	813
Final randomly selected tweets[h], n	83	100	90	65	100	100	31	569
Excluded tweets[i], n	−3	−11	−7	−4	−9	−4	0	−38
Final selected tweets, n	80	89	83	61	91	96	31	

[a]NNC: National Nutrition Committee.

[b]SSCN: Saudi Society for Clinical Nutrition.

[c]Kayl: Kayl Association for Combating Obesity.

[d]Weqaya: Public Health Authority.

[e]GHC: Gulf Health Council.

[f]SFDA: Saudi Food and Drug Administration.

[g]MOH: Ministry of Health.

[h]The process used a simple random sampling technique to select 100 tweets from each of the 7 nutritional providers' accounts in Saudi Arabia. The selection included all available tweets for accounts with fewer than 100 tweets.

[i]Duplicated, other languages, not found, or no information.

Phase 2

A simple random sampling technique was used to select 100 tweets from each of the 7 nutritional providers' accounts in Saudi Arabia. The selection included all available tweets for accounts with fewer than 100 tweets. For the 3 accounts with more than 100 tweets (SSCN, GHC, and SFDA), unique identifiers were assigned to each tweet, and a random number generator selected 100 tweets from each account. This approach produced randomly selected 569 tweets using the same random number generator to ensure unbiased representation from the initial dataset of 813 tweets.

Examination of Individual Tweets

The final 569 randomly selected diet and nutrition-related tweets were examined to collect data about the accounts and tweet characteristics. During the examination, we excluded 38 tweets due to duplication, languages, no longer available, or missing information.

A total of 531 tweets were examined and the following information was obtained: type of tweet (text, picture, video, voice, poll, and link to file), popularity (tweet likes, retweets, and number of comments), the topic of tweets (which was collected and then categorized into themes by analyzing the content of each tweet), and date and time of the tweet. From the account page on Twitter, the following information was collected: account activity or interaction with other Twitter users (following and followers), date of joining Twitter, and if there is a link to the organization. Each tweet was classified based on popularity (likes, retweets, and comments).

Expert Opinion in Nutrition

A total of 3 independent experts from clinical nutrition, food science, and technology assessed the validity, accuracy, and evidence used in the nutrition-related tweets. Tweets were extracted on a Microsoft Excel file for all text tweets; links to files (primarily guides), and polls were converted to images. Original pictures, voice, and videos were cleaned to remove colors, logos, and identifiers. All tweets were provided as an Excel folder, including all tweets in a list, a separate folder labeled by the number of tweets for each tweet with media. These were provided to independent reviewers to give their expert opinion about all 531 tweets. Reviewers were blinded to the characteristics of tweets (original tweet or retweet) and the Twitter account during content analysis. In addition, a separate sheet was provided to rate tweets accurately. Rating was performed according to the following criteria:

1. Accuracy evaluation: The reviewers classified tweets as accurate or inaccurate. The score was decided based on most reviewers' agreement [6]. If a tweet does not have a

XSL•FO
RenderX

majority score agreement, then it will be considered inconclusive accuracy.

2. Evidence-based: If accurate, did the tweet include evidence or reference, link or reliable sources, or peer-reviewed studies or officially published guidelines?

Sample Size

The sample size was estimated to find a significant difference in accuracy prevalence rate, and it was calculated based on a type one error of 0.05% and 80% power and a prevalence of outcome of 20 percent (according to Alnemer et al [6]) to be 304 tweets per awareness provider (account). Hence, it was aimed to collect a maximum of 400 most recent nutrition-related tweets by each of the 7 nutrition awareness providers in Saudi Arabia, published within 12 months for a total of 2800 tweets. A random selection of 100 tweets per account was made for the final sample to conduct the content analysis for accounts exceeding 100 tweets for validity and accuracy (the target total tweet number was 700 nutrition-related tweets). However, we could only include 531 tweets due to some accounts' small number of nutritional tweets, such as of the MOH and Weqaya.

Data Management

The collected data was analyzed using IBM SPSS (version 24.0). Nutrition-related tweets were categorized according to their accuracy level, account characteristics, and tweets. Descriptive statistics were used to tabulate frequencies and proportions, and the median of likes, comments, and retweets was computed. The chi-squared test (χ^2) was used to determine statistical significance in assessing the difference between the accuracy of nutrition-related tweets; a P value<.05 is considered significant [6].

Ethical Considerations

All Twitter user accounts and their identities were anonymized in this study. Only the affiliations of the 7 nutrition awareness providers in Saudi Arabia were identified. No personally identifiable data was included in the study. Data was stored securely in an encrypted, password-protected file. Ethical approval (IRBC/1481/21) was obtained from the Institutional Review Board office at King Abdullah International Medical Research Center (KAIMRC). This study does not require informed consent due to the public availability of the data collected from Twitter and its nature as publicly accessible information. While the data is publicly available, we ensured compliance with privacy and ethical standards by anonymizing all user information.

Results

A total of 531 tweets posted by 7 different Saudi Arabian Twitter accounts were included in the study.

Table 2 illustrates some of the main characteristics of nutrition-related tweets. Of the 531 final tweets from 7 awareness providers, 279 (52.5%) were by governmental entities, and 252 (48%) were by nongovernmental entities. The majority of tweets, 96 (18%), were extracted from SFDA's account, and the least number of tweets, 61 (12%) and 31 (5%) were extracted from Weqaya and MOH's accounts, respectively. The rest of the entity's accounts had almost the same number of tweets.

All accounts had an official website available on the Twitter page of the account except for SSCN, which accounts for 89 (17%) tweets published without an organizational website. Nearly half of the tweets, 219 (41.2%), were posted between March 2021 and May 2021, while the least number of tweets was between December 2020 and February 2021, 62 (12%).

Table 3 shows the number and proportion of tweets per month. April 2021 had the highest number of tweets, 113 (21.3%), while December 2020 and Feb 2021 had the lowest number of tweets, 18 (3%) in 12 months (see Table 3). Of the 531 tweets, 262 (49.3%) were published from 4 PM to 11:59 PM, and 233 (43.9%) were published from 8 AM to 3:59 PM. In contrast, only 36 (7%) tweets were published between 12 AM and 7:59 AM.

XSL•FO
RenderX

Table . Descriptive statistics of the nutritional tweets (N=531).

Variables	Tweets (N=531), n (%)
Account	
SFDA[a]	96 (18)
GHC[b]	91 (17)
SSCN[c]	89 (17)
Kayl[d]	83 (16)
NNC[e]	80 (15)
Weqaya[f]	61 (12)
MOH[g]	31 (5)
Type of organization	
Governmental	279 (52.5)
Nongovernmental	252 (47.5)
Organization link	
Yes	442 (83.2)
No	89 (17)
Date	
September 2020 to November 2020	142 (26.7)
December 2020 to February 2021	62 (12)
March 2021 to May 2021	219 (41.2)
June 2021 to August 2021	108 (20.3)
Time	
12 AM to 7:59 AM	36 (7)
8 AM to 3:59 PM	233 (43.9)
4 PM to 11:59 PM	262 (49.3)
Topic	
Diet or eating habits	173 (32.6)
Nutrition label, nutrients, or supplements	122 (23)
Disease or condition	109 (20.5)
Food safety	49 (9)
Ramadan	60 (11)
Water or dehydration	18 (3)
Type of content	
Text	90 (17)
Picture	342 (64.4)
Video	67 (13)
Voice	11 (2)
Poll	7 (1)
Link to file	14 (3)
Accuracy	
Accurate	445 (83.8)
Inaccurate	13 (2)
Inconclusive accuracy	73 (14)

XSL•FO
RenderX

Variables	Tweets (N=531), n (%)
Evidence-based	
With evidence	17 (4)
Without evidence	428 (96.2)

[a]SFDA: Saudi Food and Drug Administration.
[b]GHC: Gulf Health Council.
[c]SSCN: Saudi Society for Clinical Nutrition.
[d]Kayl: Kayl Association for Combating Obesity.
[e]NNC: National Nutrition Committee.
[f]Weqaya: Public Health Authority.
[g]MOH: Ministry of Health.

Table . Number and proportion of tweets by month for the 12-month period (N=531).

Month-year tweeted	n Tweets (N=531), n (%)
September 2020	64 (12)
October 2020	39 (7)
November 2020	39 (7)
December 2020	18 (3)
January 2021	26 (5)
February 2021	18 (3)
March 2021	52 (10)
April 2021	113 (21.3)
May 2021	54 (10)
June 2021	41 (8)
July 2021	42 (8)
August 2021	25 (5)

Most of the tweets' topics were about diet or eating habits, accounting for 173 (32.6%); nutrition labels, nutrients, or supplements 122 (23%), and diseases/conditions 109 (20.5%). In contrast, the least tweeted topic was about food safety at 49 (9%), Ramadan at 60 (11%), and water or dehydration at 18 (3%). Most tweets were in picture content 342 (64.4%), followed by texts 90 (17%). On the other hand, voices, polls, and links were all less than 10%. More than 3-quarters of the tweets, 445 (83.8%), were accurate, whereas 13 (2%) tweets were inaccurate, and 73 (14%) had inconclusive accuracy (no agreement) based on the expert's opinion. Among accurate tweets, 17 (4%) were evidence-based (included a mention of the evidence in the tweet), and 428 (96.2%) were not evidence-based.

The popularity of tweets based on likes, comments, retweets, and followers for each account is shown in Table 4. The median of tweet likes was 14 likes per account (range 0 - 1442). The median of retweets was 8 (range 0-352) per account, and the median of comments was 1 (range 0 - 133) per account. Finally, the median number of followers per account was 158,600 (range 1656-6,000,000).

Table 5 shows crosstabulation results describing the correlation between the accuracy of tweets and all characteristics using X^2 and fisher exact tests. There were significant associations between accuracy of tweets and the following characteristics: inclusion of evidence (P<.01, 95% CI <0.001-<0.001), source (account; P=.009, 95% CI 0.008 - 0.01), type of organization (P=.009, 95% CI 0.007 - 0.01), time tweeted (P=.04, 95% CI 0.04 - 0.04), topic of tweets (P=.003, 95% CI 0.002 - 0.004), and type of content (P=.03, 95% CI 0.03 - 0.03).

A total of 428 (80.6%) accurate tweets did not include or mention the source of evidence. Another significant relation was found between the accuracy of the tweets and the source of data (accounts). The SFDA account had the highest number of accurate tweets, 92 (17%), while MOH had the lowest number of accurate tweets, 28 (5%). The GHC had the highest number of not-accurate tweets with 5 (1%), whereas MOH had the lowest with 0 inaccurate tweets. A relationship was also found between the accuracy of tweets and the type of organization. Nongovernmental organizations had more tweets with inconclusive accuracy (n=46, 9%). There is also a relationship between accuracy and the time tweeted and between the accuracy of the tweets and the topic tweeted. Diet and eating habits had the highest number of accurate tweets, 143 (26.9%), followed by disease and conditions, 95 (18%). Yet, nutrition labels, nutrients, or supplements had the highest number of inaccurate tweets at 8 (2%). Furthermore, there was a significant association between the accuracy of the tweets and the type of published content. The most accurate tweets were pictured, 291

XSL•FO
RenderX

(54.8%). No significant differences were found between the accuracy of the tweets and the rest of the variables (organizations, link, date, number of likes, comments, and retweets).

Table 6 shows the popularity by account based on a score of the average likes, comments, and retweets and the number of account followers. The MOH had the highest rank, making it the most popular account, followed by SFDA and GHC. The NNC and SSCN had the least popularity among all accounts.

Table . Median likes, comments, retweets, and followers.

Metric	Values, median (IQR)
Likes	14 (0 - 1422)
Comments	1 (0 - 133)
Retweets	8 (0 - 352)
Followers	158,600 (1656-6,000,000)

XSL•FO
RenderX

Table . Cross-tabulation showing the correlation between the accuracy of the nutritional tweets and characteristics (N=531).

Variables	Accurate, n (%)	Inaccurate, n (%)	Inconclusive accuracy, n (%)	*P* value	95% CI
Included evidence[a]				<.001	<.001 to <.001
Yes	17 (3)	0 (0)	0 (0)		
No	428 (80.6)	0 (0)	0 (0)		
Not applicable	0 (0)	13 (2)	73 (14)		
Account[a]				.009	.008-.01
GHC[b]	72 (14)	5 (0.9)	14 (3)		
Kayl[c]	65 (12)	1 (0.2)	17 (3)		
MOH[d]	28 (5)	0 (0.0)	3 (0.6)		
NNC[e]	65 (12)	2 (0.4)	13 (2)		
SFDA[f]	92 (17)	1 (0.2)	3 (0.6)		
SSCN[g]	72 (14)	1 (0.2)	16 (3)		
Weqaya[h]	51 (10)	3 (0.6)	7 (1)		
Type of organization				.009	.007-.01
Governmental	243 (45.8)	9 (1.7)	27 (5)		
Nongovernmental	292 (38)	4 (0.8)	46 (9)		
Organization link				.34	.33-.35
Yes	373 (70.2)	12 (2.3)	57 (11)		
No	72 (14)	1 (0.2)	16 (3)		
Date[a]				.96	.95-.96
September 2020 to November 2020	117 (22)	3 (0.6)	22 (4)		
December 2020 to February 2021	52 (10)	1 (0.2)	9 (2)		
March 2021 to May 2021	185 (35)	5 (0.9)	29 (6)		
June 2021 to August 2021	91 (17)	4 (0.8)	13 (2)		
Time[a]				.04	.04-.04
12 AM to 7:59 AM	25 (4.7)	3 (0.6)	8 (2)		
8 AM to 3:59 PM	202 (38)	3 (0.6)	28 (5)		
4 PM to 11:59 PM	218 (41.1)	7 (1.3)	37 (7)		
Topic[a]				.003	.002-.004
Diet or eating Habits	143 (26.9)	4 (0.8)	26 (5)		
Nutrition labels, nutrients, or supplements	92 (17.3)	8 (1.5)	22 (4)		
Disease or condition	95 (17.9)	1 (0.2)	13 (2)		
Food safety	49 (9.2)	0 (0)	0 (0)		
Ramadan	52 (9.8)	0 (0)	8 (2)		
Water or dehydration	14 (2.6)	0 (0)	4 (1)		
Type of content[a]				.03	.03-.03
Text	64 (12.1)	3 (0.6)	23 (4)		

XSL•FO
RenderX

Variables	Accurate, n (%)	Inaccurate, n (%)	Inconclusive accuracy, n (%)	*P* value	95% CI
Picture	291 (54.8)	9 (1.7)	42 (8)		
Video	60 (11.3)	0 (0)	7 (1)		
Voice	11 (2.1)	0 (0)	0 (0)		
Poll	6 (1.1)	0 (0)	1 (0.2)		
Link to file	13 (2.4)	1 (0.2)	0 (0)		

[a]Fisher exact test.

[b]GHC: Gulf Health Council.

[c]Kayl: Kayl Association for Combating Obesity.

[d]MOH: Ministry of Health.

[e]NNC: National Nutrition Committee.

[f]SFDA: Saudi Food and Drug Administration.

[g]SSCN: Saudi Society for Clinical Nutrition.

[h]Weqaya: Public Health Authority.

Table . Popularity ranking by account score based on the average number of retweets, likes, comments, and the number of account followers (rank of 1 indicates the highest and 6 indicates the lowest popularity).

Account	Average comment		Average retweet		Average like		Follower per account		Total score of rank	Rank of popularity
	Rank	n	Rank	n	Rank	n	Rank	n		
MOH[a]	1	17.26	1	138.58	1	230.19	1	6,000,000	4	1
SFDA[b]	2	6.11	2	45.25	3	61.24	2	1,300,000	9	2
GHC[c]	3	4.88	3	32.95	2	103.60	3	294,800	11	3
Weqaya[d]	7	0.91	4	11.53	4	18.63	4	158,600	19	4
Kayl[e]	6	1.61	5	10.05	5	15.54	5	12,400	21	5
NNC[f]	5	1.76	6	7.93	7	11.39	6	5657	24	6
SSCN[g]	4	2.02	7	7.09	6	14.93	7	1656	24	6

[a]MOH: Ministry of Health.

[b]SFDA: Saudi Food and Drug Administration.

[c]GHC: Gulf Health Council.

[d]Weqaya: Public Health Authority.

[e]Kayl: Kayl Association for Combating Obesity.

[f]NNC: National Nutrition Committee.

[g]SSCN: Saudi Society for Clinical Nutrition.

Discussion

Principal Findings

This study aimed to assess the validity of nutrition-related content posted on Twitter by 7 reputable Saudi Arabian Twitter accounts that disseminate nutritional awareness information in Saudi Arabia. Both governmental and nongovernmental organizations were included in our analysis. The validity of nutrition-related tweets posted on Twitter by preidentified Saudi Arabian nutritional awareness providers was examined, and deficiencies and gaps in nutrition-related content on Twitter were identified.

This study's findings indicated that most tweets (84%) were deemed accurate based on experts' opinions. Albalawi et al [16] examined the trustworthiness of tweets and found that roughly 69% of health-related tweets were trustworthy. Yet, Alnemer et al [6] found that approximately half of the extracted health-related tweets (49%) were evidence-based. Furthermore, the findings of Lee et al [21] indicate that over half of health-related tweets (53%) contained "testable" claims supported by medical evidence. Public's tweets were less accurate than health care professional's tweets [22]. These results, both locally and internationally, are somewhat consistent with the findings of this study. Yet, the number of accurate tweets is significantly higher in this study, 84%, compared to Alnemer et al [6], Albalawi et al [16], or Lee et al [16]. This might be due to the preselection of reputable awareness providers. Therefore, the information tweeted was of higher quality. It is worth noting that all previous studies focused on health-related tweets, yet we focused on nutrition-related tweets

XSL•FO
RenderX

in this study. In addition, this study included 3 levels of accuracy (accurate, inaccurate, and inconclusive accuracy), whereas in the previous studies, they only classified tweets as accurate or inaccurate.

Among accurate tweets, only 4% of tweets cited or mentioned the source of the evidence. This might be due to the limited capacity of allowed characters per tweet. Likewise, previous studies conducted locally found that a significant number of the extracted tweets were classified as untrustworthy health information [6,16]. However, previous studies only classified tweets based on accuracy (the same as evidence-based); in this study, we also looked at including the source of evidence in the tweet to allow readers to verify or read more about the topic. Therefore, including evidence in our study has a different meaning than evidence based on previous studies [6,16,21]. Yet, this highlights that low inclusion of references in tweets is a common challenge, which might be due to the limited characters allowed on the platform. To overcome this challenge, content creators could use hyperlinks, multimedia (pictures and videos) [23], or threading tweets [24] to accommodate references without exceeding character limits.

In addition, Alassiri et al [8] reported that over half of their study population doubted the credibility of information available on Twitter, which contradicts this study's findings regarding the reliability of shared information. According to our experts, only 2% of tweets were deemed inaccurate. Again, this discrepancy might be explained by the preselection of reputable awareness providers to include in this study rather than selecting any nutritional-related tweets. Therefore, the information tweeted might be of higher quality by professionals in the field.

The study's findings also revealed that 13% of the tweets did not have a majority agreement on content accuracy by experts. Although this number is relatively small, it is an interesting finding. A possible explanation might be the influence of reviewers' backgrounds and affiliations with different schools of nutrition. Our reviewers had backgrounds in clinical nutrition, food science, and food science technology, which may have affected their evaluation of the tweets. Their differing expertise may have influenced their evaluations, highlighting the challenge of ensuring reliable nutrition content on social media. This variation in expert opinions emphasizes the need for consistent, evidence-based guidelines to improve the quality of online nutrition information. To improve consistency and reliability of the evaluation, standardized international or national nutritional guidelines could be used as a reference for evaluating content [25,26]. In addition, using consensus decision techniques such as the Delphi method could produce agreement among a group of experts in future studies [27]. This finding matches other studies in the literature; Albalawi et al [16] and Alnemer et al [6] identified conflicting opinions between study reviewers regardless of the evaluation. These findings suggest that in nutrition, conflicting opinions arise within the field of diet and nutrition, which can impact the interpretation and analysis of the tweets. This also highlights the importance of considering different perspectives and sources of information when analyzing social media content related to diet and nutrition. Therefore, it is essential to recognize the potential influence of individual biases and perspectives when analyzing social media content.

The study results indicate a significant relationship between the accuracy of tweets and the source of data (accounts). Based on expert opinions, the majority of tweets were accurate in governmental and nongovernmental organizations, with 2% inaccurate. As expected, the MOH was the most popular account, yet had the lowest number of nutritional tweets with zero inaccurate tweets. Meanwhile, the SFDA, which also represents a government organization, had the highest number of accurate tweets, followed by the GHC. Yet, nongovernmental organizations had a higher number of tweets with inconclusive accuracy. These findings were consistent with those reported by Alnemer et al [6], who found that 80% of government institute tweets were valid, and by Albalawi et al [16], who found that organizational accounts and traditionally trusted health sources were more likely to tweet accurate information. Also, the study by Lee et al [21] indicated that health care organizations such as hospitals, medical societies, health care clinics, and journals were more likely to share "testable" claims, which were expected to be supported by medical evidence, than other accounts. However, in this study, we only included reputable awareness providers who are known in health and nutrition.

The SSCN and Kayl, representing associations considered nongovernmental organizations mainly focused on nutrition, had the third and fourth highest number of accurate tweets despite being among the least popular accounts, ranked 6 and 5, respectively (see Table 5). These findings suggest that government and nongovernment organizations are reliable and trustworthy, regardless of their popularity. Although the number is small, it is notable that governmental organizations represented by the GHC had the highest number of inaccurate tweets (5, 1%), followed by Weqaya and NNC, which are governmental organizations. It is worth noting that although Weqaya had the second-lowest number of tweets related to nutrition, it had the highest number of inaccurate tweets. This highlights the possibility that even trusted health sources may publish inaccurate nutrition-related information, especially if they are nonspecialized entities in nutrition. This highlights the importance of critically evaluating nutrition information from social media platforms. While it is encouraging that government and nongovernment organizations were found to be reliable and trustworthy overall, inaccurate information underscores the need for continued efforts to ensure the accuracy and reliability of nutrition-related content on social media.

In addition, the MOH had the lowest number of nutritional tweets, which may be attributed to the relatively low number of nutrition-specific tweets posted by the account. MOH is considered one of the main awareness providers; it frequently posts about a variety of topics related to health. Its Twitter account covers a wide range of topics, including public health campaigns, disease prevention, and general health care updates especially during the pandemic. As a result, nutrition-related content may not receive as much focus or prioritization as other health topics such as COVID-19 topics. Given the timing during the COVID-19 pandemic, the MOH account focused more on evidence-based information in response to the pandemic and

XSL•FO
RenderX

increased demand for health-related content [28]. The public's heightened concern for reliable health information during the global health crisis could have influenced the nature of the tweets [28]. Also, MOH may have fewer resources or a smaller team dedicated explicitly to creating nutrition-related content compared to specialized organizations such as SFDA or NNC. Although there is limited focus on nutrition, it is suggested that MOH has a good tweet quality shared through their account. In addition, the dehydration topic had the lowest tweets, followed by food safety and Ramadan. While Ramadan-related topics are usually tweeted in the holy month, the same is not expected for dehydration and food safety. Therefore, nutritional topic priorities should be identified on a national level to be shared with awareness providers to guide creating their content and avoid underrepresentation of specific topics.

A significant association was observed between the accuracy of tweets and their topic, classified into 6 categories: diet or eating habits; nutrition label, nutrients, or supplements; disease or condition; food safety; Ramadan, and water or dehydration. Alassiri et al [8] reported that most of their study population held a favorable attitude toward using Twitter, particularly concerning nutrition and healthy lifestyles. It is encouraging to note that nutrition awareness providers in Saudi Arabia are publishing a high number of tweets related to diet or eating habits and nutrition labels, nutrients, or supplements, which are topics of particular interest to Twitter users, based on the study by Alassiri and Alowfi [8]. However, it is equally essential to ensure that information across all categories is accurate and trustworthy, as inaccurate data can seriously affect individuals and society [8].

Another significant finding was the observed association between tweet accuracy and the type of content shared. Tweets presented in picture format had a higher likelihood of being accurate compared to text-only tweets. This association might be because pictures allow for more comprehensive and visually organized information, which reduces the risk of oversimplification or misrepresentation often observed in text-constrained formats [29,30]. In addition, pictures provide an opportunity to include detailed content, such as diagrams or infographics, which can enhance accuracy by presenting evidence in a visually structured way. Twitter's character limit likely drives the preference for visual content as it enables more efficient communication [29,30]. Beyond accuracy, visual content is generally more engaging and attractive than other formats. Pictures effectively transmit information and help users understand and remember the content [29]. Furthermore, their ability to spread quickly and perform better on social media platforms enhances their reach and impact [29,30]. Pictures can also be more appealing specifically to individuals with lower health literacy [23].

No significant difference was found between the accuracy of tweets and the time they were posted. However, it is noteworthy that most of the accurate tweets with evidence were posted between 8 AM and 3:59 PM, despite the majority of tweets being posted at night. This finding is in some way matching with those of Albalawi et al [16], who found that the majority of trustworthy tweets were posted during the daytime.

No significant differences were found between the accuracy of extracted tweets and their average number of retweets and favorites. This finding is consistent with the results reported by Albalawi et al [16], although accurate tweets were more likely to be liked and retweeted by others in their study. A study conducted by Hand et al [31] discovered that the source of tweets has a significant impact on its dissemination and popularity. The study found that tweets extracted from professional, government, patient advocacy organizations, or charities were more likely to be marked as favorites and retweeted compared to tweets from other sources [31]. This finding is (in some way) in agreement with this study, which has also shown that governmental accounts tend to be more popular than nongovernmental (nonprofit) accounts. It is important to note that the study by Hand et al [31] was conducted within a specific context and within a particular location and population, which may introduce bias into the comparison.

It is important to bear in mind that the studies conducted by Albalawi et al [16], Alnemer et al [6], and Lee et al [21], all focused on the analysis of health-related tweets, while Hand et al [31] analyzed social media content related specifically to nutrition and heart failure. In contrast, this study exclusively examined nutrition-related tweets. This distinction is important to be noted when comparing the results of these studies, as the research question and the topic of analysis can significantly impact the findings.

Marar et al [14] reported that a significant number of study participants not only sought health information through social media but also found it reliable and believed it increased their health awareness and knowledge. Furthermore, over 3-quarters of participants stated that this information had an impact on their health status [14]. The findings reported here suggest that there are varying perspectives on the accuracy and reliability of nutrition or health-related content on Twitter. These highlight the importance of critically evaluating health information obtained from social media and seeking out multiple sources to ensure the information is evidence-based and reliable. Furthermore, as Marar et al [14] reported, the influence of social media on health-related knowledge and behavior is substantial, indicating the need for continued efforts to ensure the accuracy and reliability of health information disseminated through these channels. In general, these results assure the importance of promoting evidence-based nutrition information on social media and the need for consumers to approach health-related content with a critical eye.

Conclusions

In conclusion, based on our analysis, nutrition-related tweets posted by reputable awareness providers are considered mostly accurate. Yet, inclusion of the source of evidence in the tweet was not a common practice, although it may enhance the user's experience and knowledge. To enhance the reliability of nutrition content on social media, we recommend implementing citation guidelines for health-related accounts, such as providing links to credible sources and adopting recognized citation formats by using hyperlinks and multimedia. These practices

XSL•FO
RenderX

would increase transparency and trustworthiness, benefiting both content creators and users.

The results indicate that regardless of the type of organization, time, date, topic, or type of content tweeted, if the source tweeting is knowledgeable and reputable, then tweets are more likely to be accurate. Also, nutrition-related tweets have a different content type, while there is a clear preference to tweeting pictures and text. It is also observed that there are diverse topics, yet there is a clear need to balance the distribution of topics. For instance, more tweets about food safety and dehydration. These results are reassuring that social media can have accurate and evidence-based information. However, users should be selective in choosing the Twitter accounts to seek for information.

This study offers opportunities for further research, including subanalysis of accuracy levels and quality of evidence. Also, future studies could use automation tools and software such as social media APIs and machine learning to automate the identification, extraction, and analysis of tweets to improve data collection efficiency. These tools could also be used to link content creators and cross-referencing their qualifications with tweets. This will overcome the manual analysis challenges such as resource-intensive and human error.

Limitations

This study has few limitations, while the data collection started on September 2021, some of the tweets were a year old (such as posted in September 2020). This means that the landscape of tweets might have changed over time. The data collection was performed manually; although the data were screened multiple times to control for any fallibility, there is still a chance of errors and mistakes. Also, along the process of extracting tweets, some tweets were no longer available, which might have different characteristics than the included tweets. For some users (Kayl, NNC, Weqaya, and MOH), there were not enough number of tweets to randomly select; therefore, all tweets were included for these accounts. Therefore, accounts did not have an equal number of nutritional tweets to compare between.

Acknowledgments

The authors would also like to thank Dr. Yara Almuhtadi from the Department of Community Health Sciences, King Saud University (KSU), and the National Nutrition Committee (NNC); and Sarah Alkhunein from the National Nutrition Committee (NNC) in the Saudi Food and Drug Authority (SFDA) for supporting this study by creating the idea and reviewing this study. Also, we would like to thank Haya Alomran, Lana Alhumaid, Hanadi Bin Mujalli, and Aeshah Alshaya for collecting data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Extraction sheet for tweets.
[DOCX File, 460 KB - ojphi_v17i1e68128_app1.docx]

References

1. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
2. Tankovska SH. Number of social network users worldwide from 2017 to 2025. Statista. 2021 May 1. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users [accessed 2025-08-02]
3. CST issued the saudi internet report, which includes data and statistics on internet penetration and use. Communications, Space, and Technology Commission. 2023. URL: https://www.cst.gov.sa/en/business/media-center/news/CST-issued-the-Saudi-Internet-Report-which-includes-data-and-statistics-on-internet-penetration-and [accessed 2025-09-26]
4. Facebook and twitter top in number of users, over 18 million users of social media programs and applications in saudi arabia. The Ministry of Communications and Information Technology. 2021. URL: https://www.mcit.gov.sa/en/mcit-annual-report-2020 [accessed 2025-08-02]
5. Saudi Youth Statistics Report. GASTAT. 2023. URL: https://www.stats.gov.sa/documents/20117/2435273/Saudi+Youth+Statistics++en.pdf/0ad67419-e535-b6f6-4b02-6f66754fafc3?t=1742730755297 [accessed 2025-09-21]
6. Alnemer KA, Alhuzaim WM, Alnemer AA, et al. Are health-related tweets evidence based? Review and analysis of health-related tweets on Twitter. J Med Internet Res 2015 Oct 29;17(10):e246. [doi: 10.2196/jmir.4898] [Medline: 26515535]
7. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
8. Alassiri SA, Alowfi AS. Public's attitudes toward health information on twitter: a cross-sectional survey based on the Saudi Population. Cureus 2019 Oct 8;11(10):e5863. [doi: 10.7759/cureus.5863] [Medline: 31763086]
9. Dumas AA, Lapointe A, Desroches S. Users, uses, and effects of social media in dietetic practice: scoping review of the quantitative and qualitative evidence. J Med Internet Res 2018 Feb 20;20(2):e55. [doi: 10.2196/jmir.9230] [Medline: 29463487]

XSL•FO
RenderX

10. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
11. Mendoza-Herrera K, Valero-Morales I, Ocampo-Granados ME, Reyes-Morales H, Arce-Amaré F, Barquera S. An overview of social media use in the field of public health nutrition: benefits, scope, limitations, and a Latin American experience. Prev Chronic Dis 2020 Aug 6;17:E76. [doi: 10.5888/pcd17.200047] [Medline: 32762809]
12. van Dillen SME, Hiddink GJ, Koelen MA, de Graaf C, van Woerkum CMJ. Exploration of possible correlates of nutrition awareness and the relationship with nutrition-related behaviours: results of a consumer study. Public Health Nutr 2008 May;11(5):478-485. [doi: 10.1017/S1368980007000754] [Medline: 17697424]
13. Eze NM, Maduabum FO, Onyeke NG, et al. Awareness of food nutritive value and eating practices among Nigerian bank workers: implications for nutritional counseling and education. Medicine (Baltimore) 2017 Mar;96(10):e6283. [doi: 10.1097/MD.0000000000006283] [Medline: 28272248]
14. Marar SD, Al-Madaney MM, Almousawi FH. Health information on social media. SMJ 2019 Dec;40(12):1294-1298. [doi: 10.15537/smj.2019.12.24682]
15. Salzmann-Erikson M, Hiçdurmaz D. Use of social media among individuals who suffer from post-traumatic stress: a qualitative analysis of narratives. Qual Health Res 2017 Jan;27(2):285-294. [doi: 10.1177/1049732315627364] [Medline: 26825480]
16. Albalawi Y, Nikolov NS, Buckley J. Trustworthy health-related tweets on social media in Saudi Arabia: tweet metadata analysis. J Med Internet Res 2019 Oct 8;21(10):e14731. [doi: 10.2196/14731] [Medline: 31596242]
17. Chan WS, Leung AY. Use of social network sites for communication among health professionals: systematic review. J Med Internet Res 2018 Mar 28;20(3):e117. [doi: 10.2196/jmir.8382] [Medline: 29592845]
18. Alfawaz W, Alageel S, Alawdh N, Alsubaie M, Alammari D, Almuhtadi Y. Nutrition content on Twitter: the public's acceptability, use and trust. JFNR 2024 Jan 20;12(5):286-291. [doi: 10.12691/jfnr-12-5-8]
19. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187. [doi: 10.2196/17187] [Medline: 33470931]
20. Heldman AB, Schindelar J, Weaver JB. Social media engagement and public health communication: implications for public health organizations being truly "social". Public Health Rev 2013 Jun;35(1):13. [doi: 10.1007/BF03391698]
21. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. J Med Internet Res 2014 Oct 15;16(10):e237. [doi: 10.2196/jmir.3765] [Medline: 25591063]
22. Charbonneau E, Mellouli S, Chouikh A, Couture LJ, Desroches S. The information sharing behaviors of dietitians and twitter users in the nutrition and COVID-19 infodemic: content analysis study of tweets. JMIR Infodemiology 2022;2(2):e38573. [doi: 10.2196/38573] [Medline: 36188421]
23. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
24. Breu AC. From tweetstorm to tweetorials: threaded tweets as a tool for medical education and knowledge dissemination. Semin Nephrol 2020 May;40(3):273-278. [doi: 10.1016/j.semnephrol.2020.04.005] [Medline: 32560775]
25. Rycroft-Malone J. Formal consensus: the development of a national clinical guideline. Quality and Safety in Health Care 2001 Dec 1;10(4):238-244. [doi: 10.1136/qhc.0100238]
26. Singh RH, Rohr F, Splett PL. Bridging evidence and consensus methodology for inherited metabolic disorders: creating nutrition guidelines. J Eval Clin Pract 2013 Aug;19(4):584-590. [doi: 10.1111/j.1365-2753.2011.01807.x] [Medline: 22168925]
27. Brody RA, Byham-Gray L, Touger-Decker R, Passannante MR, O'Sullivan Maillet J. Identifying components of advanced-level clinical nutrition practice: a Delphi study. J Acad Nutr Diet 2012 Jun;112(6):859-869. [doi: 10.1016/j.jand.2012.02.022] [Medline: 22709813]
28. Alhassan FM, AlDossary SA. The Saudi Ministry of Health's Twitter communication strategies and public engagement during the COVID-19 pandemic: content analysis study. JMIR Public Health Surveill 2021 Jul 12;7(7):e27942. [doi: 10.2196/27942] [Medline: 34117860]
29. Liang Y, Caverlee J, Mander J. Text vs. images: on the viability of social media to assess earthquake damage. Presented at: Proceedings of the 22nd International Conference on World Wide Web; May 13-17, 2013; Rio de Janeiro, Brazil.
30. Li Y, Xie Y. Is a picture worth a thousand words? An empirical study of image content and social media engagement. Journal of Marketing Research 2020 Feb;57(1):1-19. [doi: 10.1177/0022243719881113]
31. Hand RK, Kenne D, Wolfram TM, Abram JK, Fleming M. Assessing the viability of social media for disseminating evidence-based nutrition practice guideline through content analysis of Twitter messages and health professional interviews: an observational study. J Med Internet Res 2016 Nov 15;18(11):e295. [doi: 10.2196/jmir.5811] [Medline: 27847349]

Abbreviations

GHC: Gulf Health Council

XSL•FO
RenderX

MOH: Ministry of Health
NNC: National Nutrition Committee
SFDA: Saudi Food and Drug Administration
SSCN: Saudi Society for Clinical Nutrition

Edited by E Mensah; submitted 29.10.24; peer-reviewed by A Mohammed, N Holliday; revised version received 14.03.25; accepted 14.06.25; published 26.09.25.

Please cite as:
Alammari D, Bindayel IA, Althukair N, AlRomi N, Aldubayan K, Khateeb N, Alabdrabalnabi A, Banamah B
Accuracy of Nutrition-Related Awareness Messages on Twitter (Rebranded as X) by the Nutrition Awareness Providers in the Kingdom of Saudi Arabia: Validity Content Analysis
Online J Public Health Inform 2025;17:e68128
URL: https://ojphi.jmir.org/2025/1/e68128
doi:10.2196/68128

© Duaa Alammari, Iman A Bindayel, Noura Althukair, Noura AlRomi, Khalid Aldubayan, Najla Khateeb, Ahmed Alabdrabalnabi, Banan Banamah. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 26.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Data Dashboard Acceptability, Use, and Perceived Effectiveness in Disseminating Local Overdose Data and Resources in a Rural New York State County: A Cross-Sectional Study

Zhongxuan He[1*], MPH; Monika Salvage[2], BS; Corinna A Noel[1*], PhD

[1]Department of Public and Ecosystem Health, Cornell University, 350 Tower Road, Ithaca, NY, United States

[2]Department of Mental Health, Cayuga County, Auburn, NY, United States

[*]these authors contributed equally

Corresponding Author:
Corinna A Noel, PhD
Department of Public and Ecosystem Health, Cornell University, 350 Tower Road, Ithaca, NY, United States

Abstract

Background: In 2023, Cayuga County, a rural county in New York State, developed and published a publicly available, interactive overdose dashboard highlighting demographic, geographic, and time trends in suspected overdoses as well as substance use–related resources in the community. Despite the widespread use of data dashboards in the overdose crisis, there is little evidence to suggest that these dashboards can effectively disseminate data and enable public health data-driven decision-making, especially in a rural county. We conducted an evaluation of the Cayuga County Overdose Data Dashboard to fill this knowledge gap.

Objective: Our study aimed to evaluate the Cayuga County Overdose Data Dashboard's acceptability, use, and perceived effectiveness in disseminating overdose data and resources.

Methods: Following the launch of the dashboard, an online Qualtrics survey collected feedback from individuals older than 18 years of age living or working in Cayuga County, asking respondents to reflect upon their experience using the dashboard. The 10-minute survey assessed usage patterns and motivations to access the dashboard as well as the dashboard's ease of use, most valued design features, and overall perceived effectiveness in communicating information on overdoses and local resources. Data were analyzed using descriptive statistics.

Results: From May to December 2023, a total of 61 individuals from Cayuga County completed the survey, including those with lived substance use experience (n=8, 13%) as well as their close contacts (n=28, 46%), health care providers (n=12, 20%), law enforcement (n=11, 18%), and local public health and mental health care professionals (n=27, 44%). The user-friendly design and frequent updates facilitate engagement, as 54% (n=33) of respondents reported accessing the dashboard at least monthly and 75% (n=46) using it to inform decision-making. Most thought that the dashboard was easy to use (n=59, 97%) and very effective in disseminating information (n=46, 76%). From the 8 different types of overdose-related information portrayed on the dashboard, the most valued were the locations of treatment and recovery services, scoring an average of 4.75 (SD 0.65) on a 5-point scale (1="Not important" to 5="Most important"), followed by the locations of free, publicly accessible Naloxone (mean 4.58, SD 0.89) and trends in fatal and nonfatal overdoses (mean 4.48, SD 0.81).

Conclusions: Overall, this study suggests that the Cayuga County Overdose Data Dashboard effectively disseminates information and enables data-driven decision-making in the region. When developing a community-level dashboard, our findings underscore the necessity of a user-friendly design, frequent data updates, and inclusion of key information and visuals on local overdose trends and resources.

(Online J Public Health Inform 2025;17:e68977) doi:10.2196/68977

KEYWORDS

public health data dashboard; opioid-related disorders prevention and control; data-driven decision making; data visualization; rural health; community resources

Introduction

The overdose crisis is a continuously evolving public health concern, remaining one of the leading causes of death among adults in the United States [1-3]. From 2002 to 2022, the age-adjusted rate of fatal overdoses quadrupled from 8.2 to 32.6 per 100,000 population, driven by increases in overdoses involving opioids and, more recently, fentanyl [1,4,5]. Following a 30% spike observed during the COVID-19 pandemic, fatal

XSL•FO
RenderX

overdose rates have plateaued, although deaths among certain population subgroups, like adults older than 35 years, have increased [3]. Overdoses in New York State (NYS) mirror the national trends [3]. In 2021, NYS reported 69.4 opioid-related emergency department (ED) visits and 25.3 opioid-related deaths per 100,000 residents [6]. Our study region of Cayuga County, a heavily impacted rural area in NYS, ranked in the top quartile statewide in 2021 in ED usage with 85.6 opioid-related ED visits and 26.4 opioid-related deaths per 100,000 residents [6].

As the overdose crisis continues to evolve, a localized, data-driven approach has been suggested to be effective, using real-time data to track overdose-related trends and identify high-need populations and regions to target public health interventions [7-10]. Many data dashboards have been developed with the goal of disseminating actionable data on overdoses [9,11,12]. Previous research indicates that for a dashboard to be 'actionable,' it should address a specific information need, be easy to navigate, and offer local granularity [13]. However, only a minority of public health dashboards meet these criteria [14]. Consequently, the utility of existing overdose data available to practitioners is often hindered by missing, delayed, and nongranular data, failing to provide detailed insights on high-risk populations, regions, or substances [8,10]. For example, the NYS Department of Health publishes aggregate, county-specific, and opioid-related data on a quarterly basis [15], but these data are lagged up to 6 months and contain no information on demographics, substances beyond opioids, or locations where overdoses are occurring within the county. Furthermore, while features like hotspot maps have been shown to facilitate more targeted and effective intervention strategies [9], they are often absent from data dashboards.

An internal assessment of NYS county-level dashboards revealed significant gaps in local data availability, granularity, and functionality. As of April 2024, only 15 out of 62 (24.2%) NYS counties had a publicly accessible, county-specific overdose data dashboard or similar tool. Among these, 60% included demographic information (most frequently age and gender), 40% provided geographic details on overdose locations (by zip code or hotspot maps), and only 26.7% reported suspected substances involved in overdoses. Although few (11.3%) of these dashboards are updated monthly or more frequently, they offer a timelier alternative to state-level data. These findings underscore the need for more comprehensive, specific, and frequently updated overdose data dashboards to support regional public health interventions.

Since the COVID-19 pandemic, there has been a high demand for the rapid development of public health data dashboards worldwide [13,16,17]. These tools are essential for disseminating critical public health information, yet evaluations of overdose-specific dashboards remain limited. Our work contributes to the literature in this field. While previous studies often focus purely on technical and operational aspects of public health dashboards, specific design features such as data visualizations or ease of use [16,18], there is an evidence gap regarding the ability of data dashboards to effectively disseminate information and support decision-making in the overdose crisis, especially in a rural setting [18].

To address the need for timely and granular local overdose data to inform public health action, Cayuga County developed an overdose data dashboard in early 2023. Following its launch, we conducted a comprehensive evaluation of the dashboard, collecting feedback from a diverse sample of community members, including those who work in the substance use and mental health fields. As such, our evaluation aimed to answer the following questions: (1) How is the publicly available, community-level overdose data dashboard used? (2) Is the overdose data dashboard easy to use and understand? (3) Which design features and data indicators are valued on a community-level overdose data dashboard? (4) How effective is the overdose data dashboard in disseminating overdose-related data and resources?

Methods

Overview

In collaboration with Cayuga County Mental Health and a diverse group of community partners, we developed and launched a county-specific overdose data dashboard in early 2023 using the Shiny web application framework for R (RStudio, PBC) and evaluated its acceptability, use, and perceived effectiveness in disseminating local overdose data and resources using an online survey. Data collection ran from May to December 2023.

Development of the Data Dashboard

Data Used in Dashboard

The data used in this dashboard is obtained from ODMAP (Overdose Detection Mapping Application Program). ODMAP is a web-based platform developed by the High-Intensity Drug Trafficking Area to facilitate near real-time overdose surveillance and response at the county level [19]. In Cayuga County, the Auburn Police Department, the Cayuga County Sheriff's Office, and the New York State Police provide information on each suspected overdose to which they respond. This includes the day and time, location, age, gender, and suspected substance, as well as whether the incident was fatal or nonfatal and if Naloxone was administered.

Dashboard Design

The need for a tool to disseminate granular and timely overdose–related data arose from discussions during monthly committee meetings focusing on substance use in Cayuga County. The committee, including approximately 20 volunteer representatives from county mental and public health, peer organizations, health care providers, law enforcement, community services, the court system, and higher education, collaborated with an epidemiologist (CN) from 2022 to 2023 to develop a county-specific overdose dashboard using ODMAP data. Intended end users of the dashboard included committee members, their peers, and members of the public.

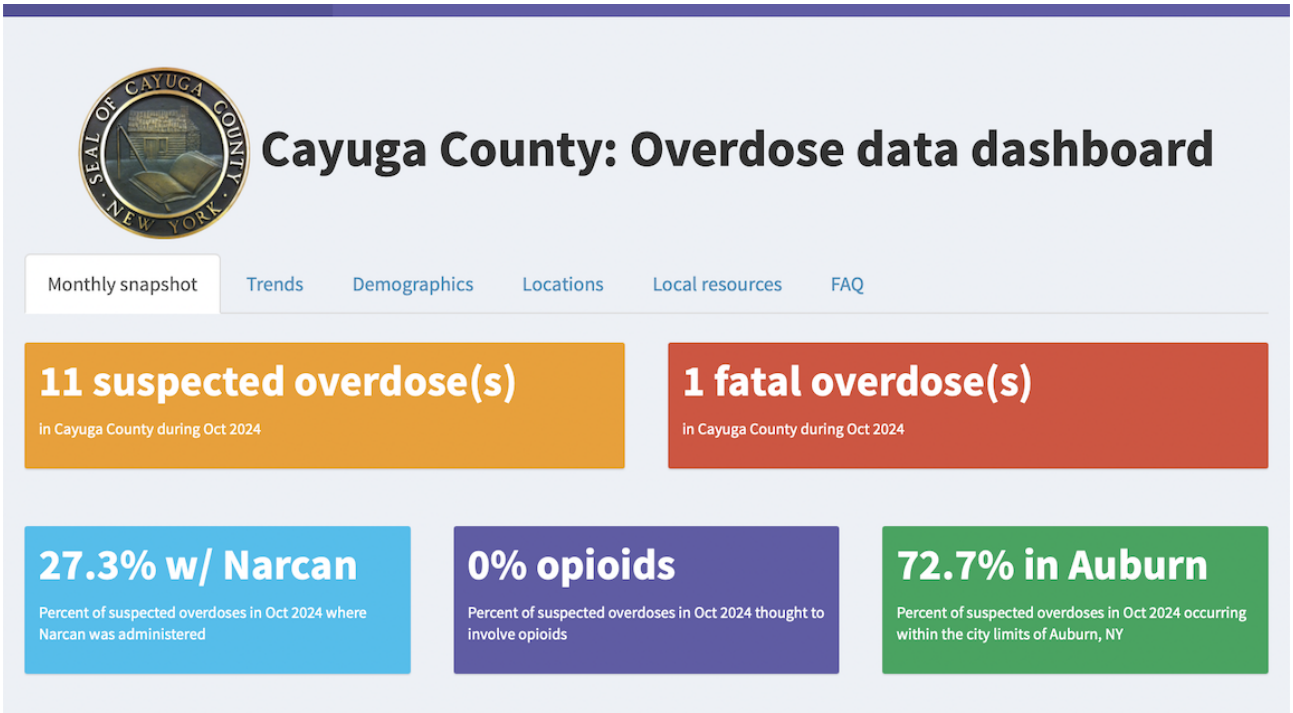
An iterative design process was taken to prioritize the ease of use, clarity, and utility of the dashboard, with the goal of enabling data-driven decision-making. The initial selection of dashboard indicators and visuals was based on previously developed static monthly reports using ODMAP data.

XSL•FO
RenderX

Committee members provided verbal and written feedback at least 4 times throughout the development process. The epidemiologist presented the most up-to-date draft dashboard in meetings and made changes based on this feedback. Comments often focused on ensuring the anonymity of those experiencing overdoses, suggesting improvements to support intended uses, and simplifying and clarifying visuals.

The final version of the dashboard was launched in March 2023 and is available on the county website. Developed using RShiny, it includes aggregate data on key overdose indicators, including a summary of the previous month, trends over the past several years (fatal and nonfatal overdoses, incidents where Naloxone was administered, and suspected substances), key demographics, and geographic locations of suspected overdoses as well as interactive features, such as customized visualizations, data filtering, and a choropleth map visualizing overdose hotspots (Figure 1 and Multimedia Appendix 1). The dashboard also provides information and maps portraying local resources, including publicly available naloxone boxes, local treatment and recovery services, and drug disposal drop boxes within the county.

Figure 1. The "monthly snapshot" tab of the Cayuga County overdose data dashboard.



Evaluation of the Data Dashboard

Following the launch of the dashboard, we conducted a cross-sectional study targeting adults older than 18 years who lived or worked in Cayuga County using a 20-question online Qualtrics survey [20].

Participant Recruitment

We conducted 2 phases of participant recruitment, resulting in a convenience sample. During the initial recruitment phase, spanning from May to August 2023, participants were offered a US $10 Amazon gift card as an incentive. They were recruited through direct links to the survey from the dashboard, a local digital newspaper segment highlighting the dashboard's launch, and a social media advertisement on county mental health accounts that included a QR code and link to the dashboard and survey. We halted the first phase of recruitment due to concerns about fraudulent responses. To filter out fraudulent responses, respondents were required to complete a supplementary verification survey to confirm the demographic information provided in the initial survey. We excluded the respondents who provided differing information on key demographics for the original and subsequent verification survey. This initial recruitment yielded 21 valid responses. The second phase of recruitment targeted community members who worked in substance use or mental health fields through email invitations to community partners and flyers with QR codes posted in public locations (county buildings and offices of substance use–related organizations) as well as an increased incentive (US $20). This approach yielded 40 more valid responses between September and December 2023. In both phases, we excluded survey responses from those who indicated that they lived outside of Cayuga County and neighboring counties.

Online Survey

The survey tool (see Multimedia Appendix 2) was designed to capture participant usage and motivations as well as perceptions of the dashboard's ease of use and effectiveness in disseminating information. First, respondents provided informed consent and were prompted to navigate to the dashboard before beginning the survey. Usage and motivations were assessed by asking respondents how, how often, and for what reasons they accessed the dashboard, how they used the information, and how frequently they thought the dashboard should be updated. Perceived effectiveness of dissemination of data and resources was evaluated using 2 items on a 5-point Likert scale (5="Very effective" to 1="Not effective"), asking respondents to rate how effective the dashboard was in: (1) "sharing data on overdoses

XSL•FO
RenderX

in the county" and (2) "sharing information on resources related to substance use in the county." Respondents also indicated whether they thought they could find and access resources using the dashboard (5="Strongly agree" to 1="Strongly disagree"). Perceived value of the types of information portrayed was assessed using a 5-point scale (5="Very important" to 1="Not important"). User experiences related to navigation and features, such as interactivity, visuals, usefulness, ability to locate the dashboard, and loading time were also evaluated on a 5-point scale (5="Strongly agree" to 1="Strongly disagree"). We used the 2-item Usability Metric for User Experience Lite (UMUX-Lite) scale as an alternative to the widely used System Usability Scale (SUS) to assess the usability and acceptability of the dashboard [21]. The UMUX-Lite has been shown to be reliable (coefficient α of 0.86), sensitive, and valid, strongly correlating with the SUS (r=0.83) and the Net Promoter Score (NPS; r=0.72) and predicting SUS scores with 99% accuracy [22,23]. The UMUX-Lite score is calculated from 2 items on a 7-point scale (7="Strongly agree" to 1="Strongly disagree"): (1) "The functionalities of the data dashboard fulfill my needs" and (2) "The data dashboard is easy to use" [22-24]. A single-item Net Promoter Score was also included to assess user satisfaction and likelihood of recommending the dashboard to others [24]. In the last part of the survey, respondents provided information on their age, gender, race/ethnicity, education, home residence, role in the community, and background related to substance use.

We implemented several security measures to maintain data integrity via the Qualtrics platform: (1) the "Prevent Ballot Box Stuffing" option to prevent multiple entries from the same individual and (2) the built-in "reCAPTCHA" and "Bot Detection" algorithms to filter out automated bot responses [20,25]. Only responses with a reCAPTCHA score of >0.5 were included.

Statistical Analysis

Descriptive statistics, including counts and percentages, and mean, SD, and range were used to summarize the study population as well as answer the primary evaluation questions. We used the regression formula proposed by Lewis et al [22] to calculate the predicted SUS from the UMUX-Lite survey items and subsequently categorized these scores based on letter grades (A-F) [22,23,26]. In exploratory analyses, we used Fisher exact tests to examine the association between the reported frequency of use of the dashboard and roles in the community, motivation for visiting the dashboard, and uses of dashboard information. Welch 2-tailed t test was used to assess the association among continuous ratings for usability, perceived value of data indicators, and user experiences related to navigation and design features, and categorical variables evaluating roles in the community, information use, and motivations. Data was analyzed in Stata (StataCorp), and a threshold of P<.05 was used to determine statistical significance [27].

Ethical Considerations

The evaluation protocol was reviewed by the Cornell University institutional review board (protocol number IRB0147362) and was deemed to be "not research" as specified by the Department of Health and Human Services Code of Federal Regulations 45 CFR 46. Informed consent was obtained for all participants.

Results

Study Population

A total of 61 people provided feedback on the dashboard through the online survey, representing a well-educated, primarily female, and White population between the age groups of 35 and 64 (mean 42, SD 4.5, range 19-70) years, with diverse roles in the community (Table 1). All participants either lived or worked in Cayuga County. Most identified as family or friends of an individual with substance use (n=28, 46%) or as working in the substance use field (n=27, 44%).

XSL•FO
RenderX

Table . Study population.

Demographic Characteristics	Values, n (%) Total (N=61)
Age (years)	
18 - 24	4 (7)
25 - 34	13 (21)
35 - 44	23 (38)
45 - 64	18 (30)
65 - 84	3 (5)
Gender	
Woman	43 (71)
Man	18 (30)
Race	
White or Caucasian	61 (100)
Education	
Some high school, but no degree	3 (45)
High school diploma or GED[a]	4 (67)
Associate degree or some college	21 (34)
Bachelor's degree	20 (33)
Postgraduate degree	13 (21)
Role in community[b]	
A person with lived experience of substance use	8 (13)
A family member/friend of someone with substance use experience	28 (46)
A health care provider	12 (20)
An individual who works with organizations that address mental health and/or substance use	27 (44)
A member of law enforcement	11 (18)
A first responder	5 (8)
A community member	26 (43)

[a]GED: General Educational Development.

[b]Total may add up to >100% because individuals could indicate more than one role.

Dashboard Use

Most respondents (n=42, 69%) had accessed the overdose dashboard prior to participating in our online survey, with 54% (n=33) visiting the dashboard monthly or more frequently, for a variety of reasons, on both mobile devices and computers (Table 2). Notably, 75% (n=46) of respondents used the dashboard to inform organizational or personal decision-making, and 44% (n=27) advocated for change related to overdose prevention and treatment. Respondents cited wanting to stay informed about overdose trends (n=44, 72%), applying it in their work (n=26, 43%), and using it to monitor the impact of overdose-related interventions (n=15, 25%). Only 8% (n=5) of respondents reported using it to find resources related to substance use.

XSL•FO
RenderX

Table . Reported dashboard usage.

Variable	Values, n (%) Total (N=61)
Dashboard visit frequency	
Daily	3 (45)
Weekly	9 (15)
Monthly	21 (34)
Yearly	9 (15)
Never	19 (31)
Device typically used to visit the dashboard	
Mobile device	16 (26)
Computer/laptop	20 (33)
Both mobile device or computer/laptop	11 (18)
Motivations for visiting dashboard[a]	
To stay informed about overdose trends in Cayuga County	44 (72)
To inform my work as a health care provider, first responder, or other community stakeholder	26 (43)
To monitor the impact of interventions aimed at reducing overdoses	15 (25)
To find resources related to substance use for myself or someone I know	5 (8)
Reported use of information on dashboard[a]	
To inform my own actions or decision-making	34 (53)
To advocate for change related to overdose prevention and treatment	28 (44)
To inform the actions or decision-making of the organization for which I work	24 (39)
To support grant writing or fundraising efforts for which organization I work	8 (13)
I don’t use the information on the dashboard for any reason	7 (12)

[a]Categories may add up to >100% because individuals could indicate more than one option.

Exploratory analyses revealed that reported visit frequency varied based on self-identified roles in the community as well as motivations for visiting the dashboard (Table 3). Community members and close contacts of individuals with substance use experience accessed the dashboard more frequently compared to other respondents (community members P=.007, close contacts: P=.03), as did those who reported wanting to stay informed about overdose trends (P=.02). Those who wished to monitor the impact of overdose-related interventions also tended to access the dashboard more frequently, although this was not statistically significant (P=.09). This contrasts with members of law enforcement, who accessed the dashboard less frequently (P<.05). Dashboard visit frequency did not vary substantially based on how participants reported using the information (P>.05).

XSL•FO
RenderX

Table . Factors influencing dashboard usage.

Factor	Dashboard usage (%)				*P* value	Values, n (%)
	Weekly or more often	Monthly	Yearly	Never		
Role in community						
a person with lived experience of substance use	50	12.5	12.5	25.0	.13	8 (13)
a family member or friend of someone with lived experience of substance use	35.7	39.3	7.1	17.9	.007	28 (46)
a health care provider	16.7	50.0	25.0	8.3	.15	12 (20)
an individual who works with organization(s) that address mental health and/or substance use	17.2	37.9	17.2	27.6	.86	29 (48)
a member of law enforcement	9.1	9.1	18.2	63.6	.05	11 (18)
a first responder	20.0	20.0	20.0	40.0	.86	5 (8)
a community member	34.6	38.5	7.7	19.2	.03	26 (43)
Motivation[a]						
To stay informed about overdose trends in Cayuga County	22.7	43.2	13.6	20.5	.02	44 (72)
To monitor the impact of interventions aimed at reducing overdoses	33.3	46.7	0.0	20.0	.09	15 (25)
To inform my work as a health care provider, first responder, or other community stakeholder	23.1	38.5	15.4	23.1	.71	26 (42.6)
To find resources related to substance use for myself	33.3	33.3	33.3	0.0	1	3 (4.92)
To find resources related to substance use for someone I know, such as a friend, family member, or colleague	50.0	50.0	0.0	0.0	.78	2 (3.28)
Use of information[a]						
To inform my own actions or decision-making	29.2	33.3	16.7	20.8	.36	24 (38.3)

XSL•FO
RenderX

Factor	Dashboard usage (%)				*P* value	Values, n (%)
	Weekly or more often	Monthly	Yearly	Never		
To inform the actions or decision-making of the organization that I work for	18.8	46.9	12.5	21.9	.16	32 (52.5)
To support grant writing or fundraising efforts for the organization I work for	37.5	25.0	12.5	25.0	.68	8 (13.1)
To advocate for change related to overdose prevention and treatment	22.2	44.4	11.1	22.2	.37	27 (44.3)
I do not use the information presented on the dashboard for any reason	—[b]	—	—	—	—	7 (11.5)

[a]Categories may add up t o>100% because individuals could indicate more than one option.

[b]Not applicable.

User Experience

The mean predicted SUS score was 79.5 (SD 8.4; range: 39.2-87.9), indicating that the Cayuga County Overdose Data Dashboard neared the eightieth percentile of SUS responses and was within the "satisfactory" range [28]. A total of 88% (n=54) of respondents rated the dashboard above a grade "B" based on the curved grading scale interpretation of the SUS score, suggesting the dashboard is perceived as a "good" product [28]. Looking at the individual items of the UMUX-Lite scale, respondents rated the dashboard as easy to use (mean 6.2, SD 0.8; range 3-7) and able to meet their needs (mean 6.3, SD 0.8; range 2-7). Those who used the dashboard to inform organizational decision-making, on average, rated the usability of the dashboard higher compared to those not using the dashboard for this purpose (*P*<.001), with a mean UMUX-Lite rating of 89.84 (SD 8.12) compared to 83.91 (SD 16.20), respectively.

The dashboard received an overall positive NPS of 21.3, with 43% (n=26) of respondents classified as promoters, 36% (n=22) as passive, and 21% (n=13) as detractors. This aligns with the predicted SUS score, suggesting that users are satisfied with the dashboard's capabilities. Despite 31% (n=19) of respondents reporting having never accessed the dashboard before participating in our survey, all (n=61, 100%) now report intending to visit the dashboard in the future.

Valued Features in Dashboard Design

Overall, the information portrayed on the dashboard and selected design features were highly rated by respondents (Table 4). Of the 8 different types of overdose-related information and resources portrayed on the dashboard, the most valued were the locations of treatment and recovery services, followed by the locations of free, publicly accessible Naloxone, and aggregate trends in fatal and nonfatal overdoses. All items were rated above a 4 on a 5-point Likert scale on average, indicating a high level of perceived importance for all. Survey respondents thought the data portrayed were useful (mean 4.3, SD 0.69; range 3-5), and were confident that they would be able to find and access naloxone (mean 4.3, SD 0.72; range 2-5), treatment and recovery services (mean 4.2, SD 0.78; range: 1-5), and medication drop boxes (mean 4.2, SD 0.78; range 2-5) in the community if needed. In open-ended feedback, some users suggested that the list of recovery resources could be expanded to include other resources beyond naloxone, medication drop-boxes, and peer, treatment, and recovery services.

Exploratory analyses suggest that close contacts of individuals with substance use experience tend to perceive suspected substances involved in overdoses as more important to portray on a dashboard compared to other respondents, with mean ratings of 4.64 (SD 0.67) and 4.30 (SD 0.68), although this was not statistically significant (*P*=.06). Mean ratings on dashboard design did not appreciably differ by reported role in the community (*P*>.05).

XSL•FO
RenderX

Table . Data indicators and design features incorporated on dashboard.

Data indicators and design features	Mean (SD) Total (N=61)
Perceived importance of information portrayed on the dashboard[a]	
Locations of treatment and recovery services	4.8 (0.65)
Locations of free naloxone	4.6 (0.89)
Trends in fatal and nonfatal overdoses	4.5 (0.81)
Trends in suspected substances involved in overdoses	4.5 (0.70)
Locations of medication drop boxes	4.4 (0.89)
Trends in naloxone administration at scenes of overdoses	4.4 (0.74)
Overdose statistics by age and gender	4.4 (0.83)
Overdose statistics of the previous month	4.4 (0.84)
Overdose statistics by geographic area	4.3 (0.80)
Dashboard design[b]	
The types of data portrayed on the dashboard are useful.	4.3 (0.69)
The interactive nature of the dashboard makes it easy for me to find the information that I am looking for.	4.3 (0.59)
Based on the information provided on the dashboard, I can easily find and access naloxone if needed.	4.3 (0.73)
The visuals have enough descriptive text for me to understand what they are showing.	4.2 (0.58)
Based on the information provided on the dashboard, I can easily find and access medication drop boxes if needed.	4.2 (0.78)
Based on the information provided on the dashboard, I can easily find and access treatment and recovery services if needed.	4.2 (0.79)
The dashboard is easy to find on the Cayuga County website.	4.1 (0.71)
The dashboard webpage takes too long to load.	2.6 (1.11)

[a]Data portrays mean rating and SD of "How important do you think it it for the dashboard to show each type of information?" on 5-point Likert scale (5=Very important to 1=Not important).

[b]Data portrays mean rating and SD of "Based on your experience using the dashboard, how much do you agree or disagree with each of the following statements?" on 5-point Likert scale (5=Strongly agree to 1=Strongly disagree).

Ability to Disseminate Information

The majority of respondents (n=46, 75%) reported that the dashboard was effective in disseminating information on both overdose statistics (mean 4.13, SD 0.90) and substance use resources (mean 4.05, SD 0.96) in the county. Ratings did not considerably differ by reported roles in the community (P>.05).

Discussion

Principal Findings

Our results suggest that the Cayuga County Overdose Data Dashboard is a valuable, effective, and useful tool for disseminating information on overdose trends and resources available in this rural community. Compared to the publicly available NYS Department of Health data, our dashboard provides more timely updates and insights into overdose trends and resources available regionally. Among the 42 respondents who used the dashboard prior to our study, 78.6% (n=33) of respondents reported accessing it monthly or more frequently, indicating consistent use in this group. Users from a wide variety of backgrounds generally report that the dashboard is easy to use and understand, with the information on the dashboard used for a variety of purposes including to inform personal or organizational decision-making, support grant writing, and/or advocate for change related to overdose prevention and treatment. In our sample, 19 out of 61 respondents had never seen the dashboard prior to completing the evaluation survey, which highlights the need for broader outreach to increase awareness and use in the community.

As overdose data dashboards become more prevalent, understanding how end users perceive and use the information is important for enhancing their public health impact. This study evaluates specific indicators, design features, and use of an overdose data dashboard, offering insights into how tools can be adapted to meet needs in similar rural settings. This work also highlights the potential of county-level data dashboards to drive action in underserved areas.

XSL•FO
RenderX

Insights and Best Practices for Developing Effective Overdose Dashboards

Selection of Indicators and Visualization of Data

Our evaluation provides valuable insights for entities wishing to develop overdose dashboards, particularly in terms of the types of information presented and how the integrated data are visualized. The Cayuga County Overdose Data Dashboard includes information on key demographics (age and gender) as well as geographic locations of overdoses and local substance use–related resources, in addition to monthly, year-to-date, and annual counts in nonfatal and fatal overdoses, naloxone administration, and suspected substances involved in overdoses. The elements included on the dashboard were identified during the development process with frequent feedback from a diverse group of community partners and are corroborated by research, which emphasizes the importance of tracking and visualizing overdose data by location, time, and population while preserving anonymity for small communities [29,30]. The Council of State and Territorial Epidemiologists has suggested best practices for visualizing jurisdictional overdose data, which include choosing the right type of chart (eg, line charts for trends and bar charts for comparisons), keeping visuals simple and clear, using a minimal color palette, and labeling all elements clearly to enhance understanding and decision-making [31]. Based on previous recommendations to highlight overdose hotspots and resources using the geographic information system [9], overdose hotspots on our dashboard are depicted with a choropleth map based on population-normalized overdose rates within town boundaries, while resources are disseminated through a map pinpointing selected resources in the community, accompanied by a list with their corresponding addresses and contact information. While some research advocates for a theory-driven approach to dashboard design [18], we opted for a less formal, community-driven approach, emerging naturally from discussions during monthly substance-use committee meetings.

Inclusion of Location-Specific Information on Resources

Our findings indicate that dashboard users value and find all the information portrayed on the dashboard useful, but especially location-specific information on substance use–related resources, which they thought they could find and access if needed, based on the information provided on the dashboard. Some users suggested that the list of recovery resources on the dashboard could be expanded to include additional services beyond naloxone, medication drop-boxes, and peer, treatment, and recovery services. This aligns with findings from other studies, which highlight that incorporating local information such as links to substance-use resources and interactive maps with features down to the neighborhood level enhances user engagement and promotes action [30,31]. Our results suggest that it is important to not only share summary statistics and trends on publicly available public health data dashboards but also actionable information on resources that community members may be able to benefit from. To be “actionable,” the resources portrayed should be relevant to users, detailed enough to allow users to find and access resources locally, and the information on resources should be easy to find and navigate to on the dashboard [13,14].

Planning for Diverse User Needs

We highlight that public health dashboards, particularly related to the overdose crisis, should plan for and accommodate diverse user needs. In our survey, 45.9% (n=28) of respondents reported multiple roles in the community, from community members to individuals working within substance use–related organizations to close contacts of people with lived experience. This highlights the need for a multifaceted and flexible dashboard interface, portraying multiple types of data to be able to meet these diverse needs. Respondents cited both personal and work-related reasons for accessing the dashboard, including a desire to be aware of the overdose trends and to inform action and decision-making around substance use. Despite these varying motivations, respondents indicated that the dashboard was able to fulfill their needs, potentially a result of early involvement of end users in the design process, as has been previously suggested [32]. Respondents also indicated that the interactive nature of the dashboard enables them to find the information they are looking for. In the Cayuga County Overdose Data Dashboard, interactivity is facilitated using dropdown menus, allowing users to filter visuals and summary statistics by date range and indicator of interest, as well as maps showing geographic locations of overdoses and resources, where users can hover their mouse to get location-specific information. This is in line with prior work, which indicates that effective data dashboards facilitate data exploration and are adaptable to a wide range of needs, workflows, and devices, stressing the importance of customizable filters that allow users to focus on the information that is most relevant to them [9,33].

Emphasis on User-Friendly Design

Our work underscores the importance of user-friendly design in promoting engagement with data dashboards. We use both the UMUX-Lite and NPS, which are well-established tools in the health care and technology fields and provide efficient and reliable measures of usability, user satisfaction, and loyalty [23,24]. To our knowledge, this is one of the first applications to use both tools to evaluate public health overdose data dashboards, reducing participant burden while providing actionable and comparable insights, making them ideal instruments for our purposes. The favorable UMUX-Lite and NPS ratings by respondents suggest that our dashboard is user-friendly and likely to be recommended to others [24]. With a predicted SUS score of 79.5, the Cayuga County Overdose Data Dashboard is perceived as more favorable when compared to other recently developed overdose data dashboards [34]. Notably, those who used the dashboard to inform organizational decision-making rated the usability of the dashboard higher than other respondents. We attribute the perception of the dashboard as user-friendly to the early and frequent feedback received from a diverse group of community partners throughout the development process [32]. As noted above, prior work suggests the use of a theory-driven approach to enhance usability, which future dashboards could consider as a guiding principle in their design process [18].

Limitations

There are several limitations to our study. With a relatively small convenience sample of 61 participants, our findings may

XSL•FO
RenderX

not represent the views of all those who live or work in Cayuga County. However, we were able to recruit a diverse sample of dashboard users from varied backgrounds. Results are only generalizable to overdose dashboards serving primarily rural regions and may not be applicable to dashboards developed for other public health problems. Given our recruitment strategy, we were unable to calculate a response rate, which introduces the risk of and our ability to assess selection bias. In addition, while we present several exploratory analyses, our study was not powered to examine differences in use, acceptability, and ability to disseminate information to individuals with different roles in the community or reasons for accessing the dashboard. Future research should examine the interplay between our outcomes and differing sociodemographics, including occupations and motivations for dashboard use. Our study focused on Cayuga County, a rural area with a predominantly White population. Due to the absence of race and ethnicity data in ODMAP, these important social constructs were not included in the dashboard or assessed in our evaluation. Future overdose data dashboards should incorporate detailed demographic data whenever possible, including race and ethnicity, to better identify disparities and guide targeted public health interventions [35,36]. In addition, while our findings show that respondents value all data indicators and rate navigation features highly, it remains unclear which specific elements in the dashboard promote action and how users leverage the information to make decisions.

Recommendations for Cayuga County

While the findings highlight the value of the Cayuga County Overdose Data Dashboard, they also reveal areas for improvement. First, outreach efforts to both the public and those working locally in the substance use and mental health fields should be increased to raise awareness and usage of the dashboard. Second, expanding the list of recovery, substance use, and social service resources could provide more comprehensive support to those that need it. Finally, incorporating more detailed demographic data may help identify disparities and guide interventions for vulnerable populations.

Conclusions

Customizable and user-friendly data dashboards such as the Cayuga County Overdose Data Dashboard represent a promising tool to effectively disseminate timely data and enable regional decision-making in the overdose crisis. These tools can support personal efforts (such as finding treatment resources, securing naloxone, or advocating for change) as well as organizational decision-making (like implementing targeted interventions), and assist in securing additional overdose-related grant funding for the region. When developing a community-level dashboard, our findings underscore the necessity of early engagement of end users as well as inclusion of aggregate statistics to describe overall trends, interactive visuals that highlight high-need populations and regions, and maps with actionable information on resources in the community.

Acknowledgments

This project was partially funded by a grant from the Bureau of Justice Administration, awarded to Cayuga County Mental Health Department.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the Cayuga County Overdose Data Dashboard.
[PDF File, 512 KB - ojphi_v17i1e68977_app1.pdf]

Multimedia Appendix 2
Survey instrument.
[PDF File, 217 KB - ojphi_v17i1e68977_app2.pdf]

References

1. Spencer MR, Garnett MF, Miniño AM. Drug overdose deaths in the United States, 2002-2022: key findings data from the national vital statistics system. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/nchs/products/databriefs/db491.htm [accessed 2024-10-11]
2. Spencer MR, Miniño AM, Warner M. Drug overdose deaths in the United States, 2001–2021. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/nchs/products/databriefs/db457.htm [accessed 2024-10-11]
3. New York State Department of Health. Opioid-related data in New York State. 2024. URL: https://www.health.ny.gov/statistics/opioid/ [accessed 2024-10-11]
4. Zoorob M. Fentanyl shock: the changing geography of overdose in the United States. Int J Drug Policy 2019 Aug;70(40-46):40-46. [doi: 10.1016/j.drugpo.2019.04.010] [Medline: 31079029]
5. Friedman J, Shover CL. Charting the fourth wave: Geographic, temporal, race/ethnicity and demographic trends in polysubstance fentanyl overdose deaths in the United States, 2010-2021. Addiction 2023 Dec;118(12):2477-2485. [doi: 10.1111/add.16318] [Medline: 37705148]

XSL•FO
RenderX

6. New York State Department of Health. New York State Opioid Annual Data Report 2023. URL: https://www.health.ny.gov/statistics/opioid/data/pdf/nys_opioid_annual_report_2023.pdf [accessed 2025-07-07]
7. Volkow ND, Chandler RK, Villani J. Need for comprehensive and timely data to address the opioid overdose epidemic without a blindfold. Addiction 2022 Aug;117(8):2132-2134. [doi: 10.1111/add.15957] [Medline: 35611646]
8. Blanco C, Wall MM, Olfson M. Data needs and models for the opioid epidemic. Mol Psychiatry 2022 Feb;27(2):787-792. [doi: 10.1038/s41380-021-01356-y] [Medline: 34716409]
9. Wiseman J. Data-driven approaches to fighting the opioid crisis. : Institute for Excellence in Government; 2019 URL: https://tinyurl.com/5x9sdp6j [accessed 2025-07-07]
10. Hoots BE. Opioid overdose surveillance: improving data to inform action. Public Health Rep 2021;136(1_suppl):5S-8S. [doi: 10.1177/00333549211020275] [Medline: 34726970]
11. Hochstatter KR, Rastogi S, Klein K, et al. Predicting accidental drug overdose as the cause of fatality in near real-time using the Suspected Potential Overdose Tracker (SPOT): public health implications. BMC Public Health 2022 Jul 8;22(1):1311. [doi: 10.1186/s12889-022-13700-0] [Medline: 35804334]
12. Wu E, Villani J, Davis A, et al. Community dashboards to support data-informed decision-making in the HEALing communities study. Drug Alcohol Depend 2020 Dec 1;217(217):108331. [doi: 10.1016/j.drugalcdep.2020.108331] [Medline: 33070058]
13. Sorapure M. User perceptions of actionability in data dashboards. J Bus Tech Commun 2023 Jul;37(3):253-280. [doi: 10.1177/10506519231161611]
14. Ivanković D, Barbazza E, Bos V, et al. Features constituting actionable COVID-19 dashboards: descriptive assessment and expert appraisal of 158 public web-based COVID-19 dashboards. J Med Internet Res 2021 Feb 24;23(2):e25682. [doi: 10.2196/25682] [Medline: 33577467]
15. New York State County opioid quarterly reports. New York State Department of Health. 2014. URL: https://apps.health.ny.gov/public/tabvis/PHIG_Public/opioid-quarterly/reports/#state [accessed 2025-02-27]
16. Narayan KA, Nayak M. Need for interactive data visualization in public health practice: examples from India. Int J Prev Med 2021;12(16):16. [doi: 10.4103/ijpvm.IJPVM_171_20] [Medline: 34084313]
17. Salvi A, Gillenwater LA, Cockrum BP, et al. Development of a real-time dashboard for overdose touchpoints: user-centered design approach. JMIR Hum Factors 2024 Jun 11;11:e57239. [doi: 10.2196/57239] [Medline: 38861717]
18. Schulze A, Brand F, Geppert J, et al. Digital dashboards visualizing public health data: a systematic review. Front Public Health 2023;11(11):999958. [doi: 10.3389/fpubh.2023.999958] [Medline: 37213621]
19. Delcher C, Horne N, McDonnell C, et al. Overdose detection mapping application program expansion evaluation—A qualitative study. Criminol Public Policy 2023 Aug;22(3):491-516 [FREE Full text] [doi: 10.1111/1745-9133.12628]
20. Qualtrics. Qualtrics. Version used May 2020. 2020. URL: https://www.qualtrics.com [accessed 2023-12-20]
21. Lewis JR JR, Utesch BS, Maher DE. UMUX-LITE: when there's no time for the SUS. 2013 Apr Presented at: Proceedings of the SIGCHI conference on human factors in computing systems; Apr 27 to May 2, 2013; Paris p. 2099-2102. [doi: 10.1145/2470654.2481287]
22. Lewis JR, Utesch BS, Maher DE. Investigating the correspondence between UMUX-LITE and SUS scores. In: Marcus A, editor. Design, User Experience, and Usability: Design Discourse Lecture Notes in Computer Science: Springer; 2015:204-211 URL: https://link.springer.com/chapter/10.1007/978-3-319-20886-2_20#citeas [accessed 2025-06-04]
23. Borsci S, Buckle P, Walne S. Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? Appl Ergon 2020 Apr;84:103007. [doi: 10.1016/j.apergo.2019.103007] [Medline: 31785449]
24. Sasmito GW, Zulfiqar LOM, Nishom M. Usability testing based on system usability scale and net promoter score. 2020 Presented at: 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI; Dec 5-6, 2019; Yogyakarta, Indonesia p. 540-545. [doi: 10.1109/ISRITI48646.2019.9034666]
25. Fraud detection. Qualtrics. URL: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection [accessed 2023-05-02]
26. Borsci S, Federici S, Bacci S, et al. Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. Int J Hum Comput Interact 2015 Aug 3;31(8):484-495. [doi: 10.1080/10447318.2015.1064648]
27. StataCorp. Stata statistical software: release 18. : StataCorp LLC; 2023.
28. Sauro J, Lewis JR. Quantifying the User Experience, 2nd edition: Morgan Kaufmann; 2016. [doi: 10.1016/B978-0-12-802308-2.00002-3]
29. Fliss MD, Cox ME, Proescholdbell S, et al. Tying overdose data to action: North Carolina's opioid and substance use action plan data dashboard. J Public Health Manag Pract 2023;29(6):831-834. [doi: 10.1097/PHH.0000000000001796] [Medline: 37498535]
30. Yedinak J, Krieger MS, Joseph R, et al. Public health dashboards in overdose prevention: the Rhode Island approach to public health data literacy, partnerships, and action. J Med Internet Res 2024 Feb 12;26:e51671. [doi: 10.2196/51671] [Medline: 38345849]

XSL•FO
RenderX

31. Council of State and Territorial Epidemiologists (CTSE). Best practices for data visualization. : CSTE Public Health Drug Overdose Surveillance Training Series for Local/Territorial Jurisdictions; 2021 URL: https://tinyurl.com/j7pyyd3y [accessed 2025-03-08]
32. Krackov A, Martin EG. Visualizing the opioid epidemic: lessons learned from a dozen years of engaging communities with data. J Public Health Manag Pract 2019;25(4):398-401. [doi: 10.1097/PHH.0000000000001037] [Medline: 31136514]
33. Franklin A, Gantela S, Shifarraw S, et al. Dashboard visualizations: supporting real-time throughput decision-making. J Biomed Inform 2017 Jul;71:211-221. [doi: 10.1016/j.jbi.2017.05.024] [Medline: 28579532]
34. Fareed N, Olvera RG, Wang Y, et al. Lessons learned from developing dashboards to support decision-making for community opioid response by community stakeholders: Mixed methods and multisite study. JMIR Hum Factors 2024 Sep 9;11:e51525. [doi: 10.2196/51525] [Medline: 39250216]
35. Kapadia F. Data democratization for health equity: a public health of consequence, August 2023. Am J Public Health 2023 Aug;113(8):839-840. [doi: 10.2105/AJPH.2023.307350] [Medline: 37406266]
36. Thorpe LE, Gourevitch MN. Data dashboards for advancing health and equity: proving their promise? Am J Public Health 2022 Jun;112(6):889-892. [doi: 10.2105/AJPH.2022.306847] [Medline: 35446603]

Abbreviations

ED: emergency department
NPS: Net Promoter Scale
NYS: New York State
ODMAP: Overdose Detection Mapping Application Program
SUS: System Usability Scale
UMUX-Lite: Usability Metric for User Experience-Lite

Edited by E Mensah; submitted 19.11.24; peer-reviewed by AL Dir, BR Harris, J Yedinak; revised version received 08.03.25; accepted 08.05.25; published 10.07.25.

Please cite as:
He Z, Salvage M, Noel CA
Data Dashboard Acceptability, Use, and Perceived Effectiveness in Disseminating Local Overdose Data and Resources in a Rural New York State County: A Cross-Sectional Study
Online J Public Health Inform 2025;17:e68977
URL: https://ojphi.jmir.org/2025/1/e68977
doi:10.2196/68977

© Zhongxuan He, Monika Salvage, Corinna A Noel. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 10.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Posttraumatic Stress Disorder Content on TikTok: Cross-Sectional Analysis of Popular #PTSD Posts

Brittany Rohl[1*], PhD; Laura Carolyn Jones[1*], MS; Rachel Nattis[1], BS; Robert Dale Claar[1], BS; Xavier Velez[1], BS; Joy Gabrielli[1], PhD; John Williamson[2], PhD; Eric Porges[1], PhD

[1]Clinical and Health Psychology, College of Public Health and Health Professions, University of Florida, 1225 Center Drive, Gainesville, FL, United States

[2]Department of Psychiatry, College of Medicine, University of Florida, Gainesville, FL, United States

[*]these authors contributed equally

Corresponding Author:
Brittany Rohl, PhD
Clinical and Health Psychology, College of Public Health and Health Professions, University of Florida, 1225 Center Drive, Gainesville, FL, United States

Abstract

Background: TikTok became an increasingly popular platform for mental health discussions during a major global stressor (COVID-19 pandemic). On TikTok, content assumed to promote user engagement is delivered in a hyperindividually curated manner through a proprietary algorithm. Mental health providers have raised concerns about TikTok's potential role in promoting inaccurate self-diagnoses, pathologizing normal behaviors, and fostering new-onset symptoms after exposure to illness-related content, such as tic-like movements linked to conversion or factitious disorders. The accuracy of posttraumatic stress disorder (PTSD)–related content with respect to conveying symptoms, diagnosis, and treatment deserves further investigation.

Objective: This study aimed to characterize the accuracy of PTSD-related TikTok content.

Methods: In February 2022, a search was conducted on TikTok using the hashtag #PTSD, and the 100 most-liked videos were retrieved. Videos were excluded if they were in a non-English language, duplicated, unrelated to PTSD, lacked audio, or contained nonfunctioning links. A publicly available Python package (TikTokPy) was used to scrape available metadata (views, shares, etc). Using the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-AV), videos were independently coded by 2 reviewers for the overall accuracy of the video (useful, personal experience, or misleading), whether the creator self-identified as a health care professional, symptoms mentioned, and overall video understandability and actionability. A third reviewer was consulted in the rare instances of coding disagreements.

Results: Of the 100 included videos, 29 were classified as useful, 59 were classified as personal experience (subjective experience without outright inaccuracies), and 12 were classified as misleading. The degree to which PTSD-related information was accurate was not associated with its understandability, actionability, or user engagement. Besides useful videos being longer (mean 88.7, SD 63.1 seconds) than personal experience videos (mean 42.7, SD 44.5 seconds), no group differences in video metadata were observed across the number of views, likes, shares, or comments ($P>.05$). While self-identified HCPs were more likely to post useful content, they also contributed to 33% (4/12) of misleading videos. Changes in cognition and mood were the most frequently reported symptoms of PTSD (38/100, 38% of videos).

Conclusions: Our findings were roughly consistent with previous studies of mental health–related TikTok content accuracy, although this is variable by diagnosis. TikTok's continuously adaptive algorithmic content delivery may expose users to nonspecific and potentially misleading "click-bait" mental health information, which could influence symptom interpretation and clinical presentation. Clinicians should be aware of the digital landscape shaping patients' perceptions of PTSD.

Trial Registration: OSF Registries osf.io/vsr4t; https://osf.io/vsr4t

(Online J Public Health Inform 2025;17:e71209) doi:10.2196/71209

KEYWORDS

posttraumatic stress disorder; TikTok; social media; health communication; PTSD

Introduction

TikTok (ByteDance) is a short-form, video-based social media platform that surged in popularity during the COVID-19 pandemic and currently has more than 1 billion active users worldwide. TikTok joins a long history of internet-based platforms that serve as media for sharing health-related information. Preliminary work describing content related to

XSL•FO
RenderX

posttraumatic stress disorder (PTSD) on TikTok suggests that there are 2 primary themes: instructional videos (21.6%) and the sharing of personal stories (67.3%) [1]. The authors cautioned that in the context of minimal trigger warnings on potentially upsetting and harmful first-person content, PTSD-related videos on TikTok may pose a risk for health misinformation and possibly vicarious traumatization. The aim of this study was to report the first systematic assessment of information quality within PTSD-related TikTok content.

The format of TikTok's platform may make it particularly vulnerable to the spread of misinformation. Content is delivered to each user's homepage ("For You Page" or FYP) in an algorithmically determined, hyperindividually curated manner, with nearly all delivered content having been produced by relative strangers. This contrasts with other social media platforms in which users tend to play a more active role in curating their content feeds (ie, by choosing profiles to "friend" or "follow"). Although the precise algorithms used to drive content delivery on TikTok are proprietary and not publicly known, they are assumed to maximize the likelihood of user engagement (time spent viewing each video, "liking" and sharing content, and conversion of exposure to a "following"). These metrics are assumed to shape subsequent content delivery with each swipe from video to video, resulting in the experience of being funneled into easily permeable communities that appear to cluster around shared ideas, experiences, traits, and observations, among other characteristics [2]. As the application and its algorithms "learn" about the user's preferences, it can come to represent an "algorithmized self," in which individual characteristics are identified without the person having ever searched for related terms [3,4].

This landscape poses two major challenges to accurate dissemination of mental health–related content specifically: (1) the sensitivity and specificity with which niche content is delivered to appropriate users, and (2) the degree to which delivered content reflects scientific consensus about particular disorders and their treatment. For example, it is plausible that the algorithm could identify a user's generalized distress, offer content related to a handful of psychological phenomena, and derive future content delivery via algorithmic personalization based on the characteristics of their engagement. Assuming some variability in the reasons why someone might engage with a piece of content, engagement and subsequent narrowing of algorithmically derived content feeds may not necessarily correspond to the accuracy, applicability, or appropriateness of the content for the particular user. Importantly, this raises questions about whether vulnerable individuals are at particular risk of misinformation.

Three authors of this paper (BR, LCJ, and RN) observed that symptom misattribution is particularly fraught with respect to posttraumatic stress, often termed "trauma responses" in online discourse. According to publicly available Google Trends data [5], searches for the phrase "trauma response" nearly doubled in late 2019, more than tripled by August 2020, and increased 900% by Fall 2022 compared to pre-2019 averages. Similar but relatively smaller increases were observed in searches for the term "trauma," while longtime upward trends in searches for "PTSD" maintained their relative trajectory since such data began collection in 2004. The proliferation of media to meet the demand for trauma-related information reflected both factual information as well as misinformation, including the misclassification of vague and banal phenomena as related to trauma. A 2021 article in the popular internet-based magazine *Slate* reports, "The trend of trauma-ifying common behaviors is so pervasive that there are now viral jokes about it… …[the trend] is relaying a vast assortment of relatable annoyances and pitfalls of being human while simultaneously upping the stakes by saying, 'We're all this way because we are traumatized'" [6]. An epitomical TikTok by user @thebrrashmoron describes it best: "Hitting snooze more than once is a sign of four mental illnesses! I heard that from a TikToker with no degree or qualifications" [7].

Clinical mental health providers have raised concerns that TikTok use may be associated with inaccurate self-diagnoses, pathologization of normal experiences, and the propagation of new-onset symptoms seemingly acquired after viewing illness-related content (particularly, tic-like movements), which may represent sociogenic expressions of conversion or factitious disorders [8,9]. A study of TikTok content related to attention-deficit hyperactivity disorder (ADHD) reported that, of the most-popular videos using the hashtag #ADHD, 52% were classified as misleading with respect to symptomatology, diagnosis, or treatment, defined as being beyond typical variability in subjective patient experience, and involving factually incorrect information [10]. Examples of misleading content included assertions that "anxiety shivers," "random noise-making," and a "lack of object permanence" are "little known" symptoms of ADHD.

This study aims to assess the accuracy of TikTok content classified under the hashtag #PTSD. We chose #PTSD for several reasons. The first was due to the relative ease with which we could verify claims (ie, comparison to diagnostic criteria and studies using diagnostically confirmed patient groups). Second, the use of #PTSD rather than #traumaresponse allowed for clear expectations with respect to the timeframe of symptom onset and what constitutes a traumatic event. This may reflect a more clinically useful framework if attempting to generalize results to patient populations. Finally, the classification of symptoms across a threshold of "legitimacy" as a means of defining group membership may be an important component of online behavior in patients with posttraumatic stress [11].

We hypothesized that user engagement would favor unique or highly salient information, such that videos classified as misleading or as reflecting subjective experiences—not incorrect but based on individual perspectives—would receive higher engagement across all measures (likes, views, shares, and comments) compared to videos classified as wholly accurate. Due to perceptions of credibility, we hypothesized that videos posted by self-identified health care professionals (HCP) would receive higher engagement across all metrics as compared to creators who did not self-identify as HCPs (non-HCP). Finally, we hypothesized that information usefulness and HCP status would predict the degree to which information was effectively communicated.

XSL•FO
RenderX

Methods

Overview

This study was preregistered as an exploratory observational investigation with Open Science Framework.

Sample Size

The sample size of 100 TikTok videos was chosen as a feasible and representative sample size to capture the most viewed videos on the platform. Previous studies have used similar sample sizes to analyze health-related videos on social media platforms such as TikTok and YouTube [10,12-15].

Search and Selection Procedures

On February 20, 2023, a new TikTok account was created for the purpose of this study using the desktop version of TikTok. The use of this study-specific account (occurring off-campus and using a nonuniversity-owned computer) was intended to avoid the influence of individual user data on the video retrieval algorithm. New TikTok accounts are prompted to choose a number of general interest categories to determine initial homepage offerings; however, we circumvented these influences by opting to skip this step. On the same day the account was created, the search function was queried with the hashtag "#PTSD." On TikTok, as on many internet platforms, hashtags serve to both categorize videos and to assist algorithms in determining content that might be of interest to users [16].

To provide a buffer in anticipation of excluding some of the retrieved videos based on eligibility criteria, the first 200 results under the "Top" tab (suggesting that results are most-suggested by TikTok's proprietary algorithm) were captured as hyperlinks, and 2 authors with expertise in Python (RDC and XV) applied an open-source data scraper called TikTokPy [17] to collect video metadata (ie, number views, likes, and shares). A similar approach was used in a previous study of the accuracy of health-related information on TikTok [10]. The Python package used in this previous study (TikTok Scraper [18]) was attempted but did not work, likely owing to coding changes in TikTok itself. The resulting dataset was sorted in descending order by view count, and inclusion and exclusion criteria were applied until 100 videos were determined to be eligible for further analysis.

Videos were included if they mentioned or implied content related to PTSD, including but not limited to etiology, symptomatology, diagnosis, lived experience with PTSD, witnessing PTSD symptoms in others, or PTSD management strategies. Exclusion criteria included non-English videos (n=0), videos completely unrelated to PTSD (n=10), duplicate videos (n=1), videos without audio (n=3), or those with nonfunctioning links suggesting that the video had been removed from the platform (n=5). A total of 19 of the initial 200 videos were excluded before the target sample of 100 was reached.

Rater-Coded Variables

Two raters with clinical research experience related to PTSD (LCJ and RN) independently assessed all 100 videos using the Patient Education Materials Assessment Tool for Audio Visual Materials (PEMAT-AV) [19]. The PEMAT-AV is a reliable tool designed to evaluate the clarity and effectiveness of audio-visual patient education videos and is designed to assess their comprehensibility and usefulness. Of note, the PEMAT-AV has traditionally been applied to patient education materials developed by health professionals. Therefore, to help contextualize ratings, support internal consistency by assessors, and in line with other studies applying this measure to social media videos, raters took notes about each video's content and style [10,20,21]. These notes were not systematically analyzed for this study but select descriptions are provided for contextual purposes. The PEMAT-AV instrument generates an overall total score as well as subscales pertinent to understandability and actionability, all expressed as percentages of total possible points. Per PEMAT-AV rating protocol, understandability was scored based on the clarity of the content's purpose, word choice and style, organization, design, and use of visual aids. Actionability was scored based on whether the content identifies actions the user can take, breaks down any actions into manageable and explicit steps, and addresses the user directly when describing actions.

Consistent with previous studies of health information on TikTok, additional analysis included classification of video content into one of the following categories: (1) useful, (2) personal experience, or (3) misleading. Useful videos presented entirely accurate information about PTSD prevention, diagnosis, prevention, prognosis, symptoms, and treatments. It is important to note that descriptions of nonspecific symptoms, as long as they are not inconsistent with PTSD, were not necessarily grounds for removal from this category. Personal experience videos presented anecdotal information related to PTSD diagnosis, etiology, symptoms, or treatment that was not inconsistent with diagnostic criteria, research, or generally accepted clinical practice. Misleading videos contained speculative information that was uncorroborated by diagnostic criteria, research, or generally accepted clinical practice. Importantly, videos that were mostly factual or focused on personal experiences could be classified as misleading if they contained any factually incorrect information. Where appropriate, the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) criteria for PTSD were applied to determine the consistency of presented symptoms with diagnostic criteria [22]. In the rare event of a disagreement, a third rater (BR) resolved disagreements between the other two raters, which occurred rarely. Representation of trauma types, *DSM-5* diagnostic criteria, and the video's tone (meme or joke, educational, personal, or myth-debunking) were also coded.

Statistical Analysis

Extreme values in metadata were expected given that the phenomenon of "virality" among highly engaged social media content is defined as substantially higher engagement than "nonviral" content. While extreme values in metadata may not represent "average" PTSD-related TikTok content, exceptionally high engagement offers important clues about its relevance to users, their reactions, and the degree of dissemination. Therefore, videos determined to be exceptional on any user engagement metric (3 or more SDs above or below the overall mean were removed for separate qualitative analysis; n=3). Resulting distributions were converted to *z* scores for ease of

XSL•FO
RenderX

comparison across metrics but remained highly nonnormal (Table 1). Therefore, incremental power-based transformations were applied until normality was achieved, with all requiring Blom transformation [23]. Previous studies assessing highly skewed distributions of social media metadata have also used power-based transformations that preserve rank order [24]. As a result of these transformations, group analyses are interpreted with respect to the significance and magnitude of difference, but not absolute difference between the original metrics.

Table . Top 100 #PTSD TikTok video characteristics by usefulness group (*z* scores; N=100).

Characteristics	Misleading (n=12), mean (SD)	Personal experience (n=59), mean (SD)	Useful (n=29), mean (SD)	*P* value
Video length	0.13 (0.83)	−0.21 (1.0)	0.71 (0.87)	.01[a]
Views	−0.04 (0.82)	0.10 (1.10)	−0.38 (0.70)	.17
Likes	−0.07 (0.84)	0.14 (1.07)	−0.48 (0.79)	.13
Shares	−0.09 (0.80)	0.06 (1.04)	−0.06 (1.21)	.80
Comments	−0.09 (0.91)	0.08 (1.0)	−0.16 (1.14)	.63

[a]*P*<.05

To compare groups defined by usefulness ratings (useful, personal experience, and misleading) across variables measuring user engagement (likes, shares, and comments), one-way ANOVAs were conducted, with Welch F correction applied when assumptions of heterogeneity were not met. In the cases where a significant group effect was observed, Student *t* post hoc tests were used, unless nonhomogeneity of variance required the use of nonparametric Games-Howell post hoc test. Nonparametric Kruskal-Wallis tests were used to assess PEMAT-AV understandability and total scores due to nonnormality. A chi-square test was used to examine PEMAT-AV actionability as a dichotomous variable (actionable vs not) due to a high number of 0 values.

Ethical Considerations

This study was submitted to the University of Florida Institutional Review Board, which determined it to be exempt from full board review because all information was obtained from publicly available content. Given that only publicly available data were collected and analyzed, there was no need to safeguard participant information.

Results

Video Quality Classification

Of the 100 included videos, 29 were classified as useful, 59 were classified as personal experience (subjective experience without outright inaccuracies), and 12 were classified as misleading. Examples of content related to each category are described in Textbox 1. Figure 1 outlines the procedures used for search, selection, and rater coding.

XSL•FO
RenderX

Textbox 1. Descriptions of example #PTSD TikTok videos by usefulness classification.

Useful (n=12)

- An infographic containing two pie charts is superimposed over an image of a sunset, and lo-fi music is playing. One pie chart is titled, "What people think PTSD is," and consists of one category labeled, "Not being able to move on after a traumatic event." The other pie chart is titled, "What PTSD actually is" and is comprised of a number of equally proportioned categories labeled as "unwanted memories," "negative self-image," "hypervigilance," "emotional distress," "sense of distrust," "intrusive thoughts," "sense of threat," "avoidance/isolation," "memory problems," "anger, guilt and shame," "anxiety/depression," "excessive blame," "dissociation," "easily scared," "flashbacks," "sleeping problems," and "self-destructive behaviors."
- A self-identified clinical psychologist explains the differences in diagnostic criteria between posttraumatic stress disorder (PTSD) and borderline personality disorder.

Personal experiences (n=59)

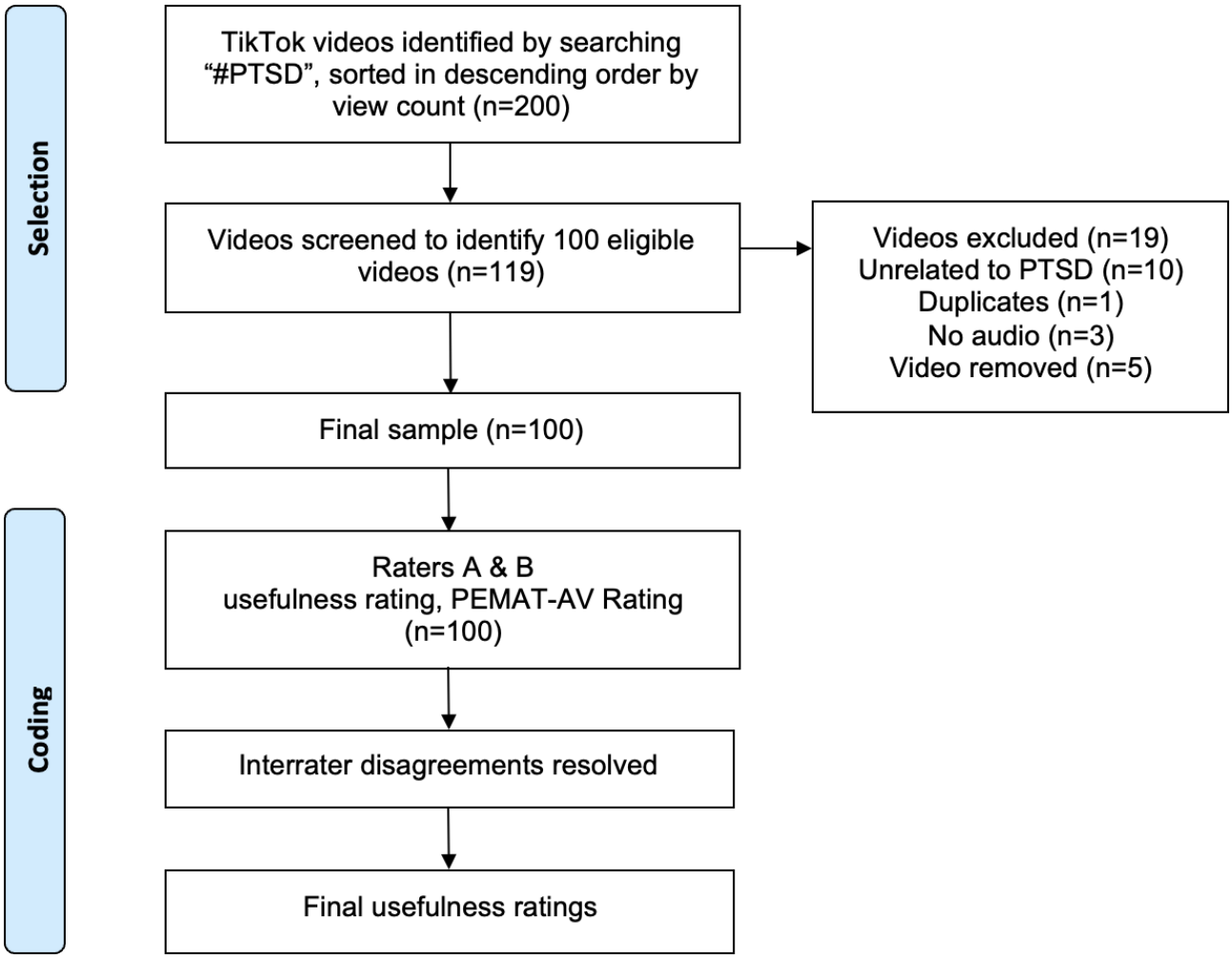
- A selfie-style video of a man saying the following: "Three things you should not do to a Veteran with PTSD: 1) Walk up behind them. You're going to end up startling them, and you're going to end up getting hurt. And it doesn't matter who you are. 2) Scare pranks. You may just have the same result as the first one we just talked about, which, you're going to get hurt. Second because some of us take hours to get our head straight for the day, and being scared like that puts us back into that state of mind that we work so hard to get out of. 3) Ask about combat experiences. If we haven't told you already about our military experiences and what we experienced while we were deployed, the chances of us telling you are slim to none. Mainly because we don't want to remember. A lot of us lost friends and family. If we do decide to share, nine times out of ten we put ourselves into a thousand-yard stare where we're re-living the situation all over again. The more you know."
- A woman films herself waking up and having "bad anxiety" which quickly triggers her PTSD. A "flashback" is preceded by eye twitching. She states that she is grounding herself in reality by touching the counter and looking around for visible proof that she is not physically back at the scene of a trauma.

Misleading (n=29)

- Clips from video games are accompanied by sound effects associated with each game. Superimposed text reads, "Sounds that will give gamers PTSD." As the game-specific sound effects play, superimposed text denotes the name of each video game.
- A self-identified clinical psychologist discusses "what people don't understand about complex trauma." While some accurate information is presented, inaccuracies include "sensitivities to sounds like breathing and chewing" and "maladaptive daydreaming" with a "rich fantasy life."

XSL•FO
RenderX

Figure 1. Outline of search, selection, and rater coding procedures for top 100 #PTSD TikTok videos. Video selection was completed by author BR and coding was completed independently by authors LCJ and RN. In the rare event of interrater disagreements, BR provided resolution. PEMAT-AV: Patient Education Materials Assessment Tool for Audiovisual Materials; PTSD: posttraumatic stress disorder.



Characteristics of Outliers

A total of 3 videos were removed from primary analyses due to extreme metadata values. Of these, one (classified as personal experience) exhibited extreme values across all metadata variables, having earned 38.6 million views (7.25 SDs above mean), 72 million likes (8.2 SDs above mean), 44,000 shares (3.6 SDs above mean), and 26,100 comments (3.58 SDs above mean). Qualitative notes indicate that the video contains a depiction of a service animal helping a trauma survivor cope with navigating the physical site where the event occurred. The other 2 videos varied with respect to which variable was extreme, with one having only a high number of comments (11.8 million, 7.25 SDs above mean, personal experience, content depicts a personal trauma narrative), and the other having only a high number of shares (83,300, 7.3 SDs above mean, misleading, content depicts a joke).

Video Metadata and Usefulness Group

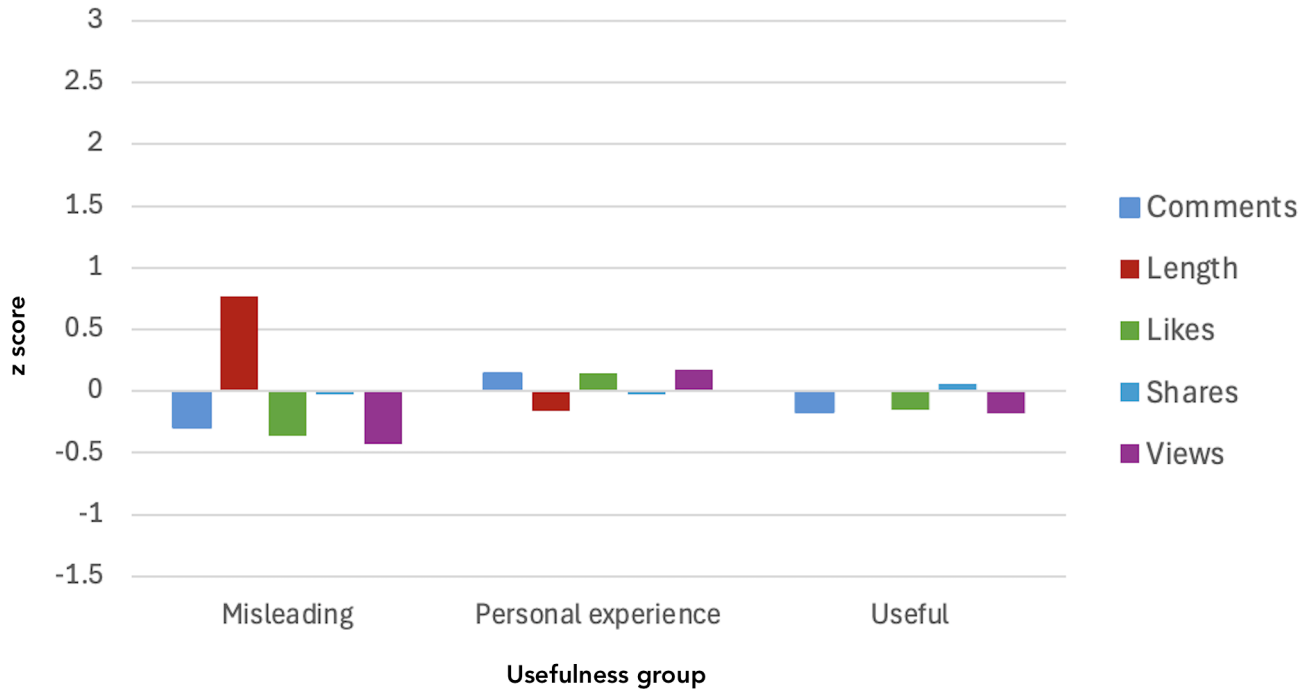
Results are presented in Table 2 and Figure 2. A significant group difference was observed for video length ($F_{2,30.7}$=5.34; P=.01) such that useful videos tended to be longer (mean 88.7, SD 63.1 seconds) than personal experience videos (mean 42.7, SD 44.5 seconds). No group differences were observed across the number of views (χ^2_2=4.42; P=.11), likes (χ^2_2=5.19; P=.07), shares (χ^2_2=0.94; P=.63), or comments (χ^2_2=2.45; P=.29)

XSL•FO
RenderX

Table . Raw top 100 #PTSD TikTok videos' metadata.

Characteristics	Misleading (n=29)	Personal experience (n=59)	Useful (n=12)	Overall group (N=100)
Video length (seconds), mean (SD)	50.28 (48.91)	42.69 (44.45)	88.67 (63.14)	50.41 (49.9)
Views, mean (SD; range)	1,981,552 (2,823,886; 25,100-11,200,000)	3,716,873 (5,947,155; 12,800-38,600,000)	737,475 (811,699; 33800-2,500,000)	2856102 (4925371; 12800-38600000)
Likes, mean (SD; range)	273,956 (423,923; 2,251-1,500,000)	520,741 (1,017,290; 1867-579,100)	99,070 (160,473; 1,867-579,100)	398,572 (827,222; 801-7,200,000)
Shares, mean (SD; range)	5374 (16,265; 125-83,300)	4457 (7384; 9-44,000)	4391 (8,355; 43-29,600)	4714 (10,709; 9-83300)
Comments, mean (SD; range)	2266 (2932; 1-12,000)	4244 (7862; 27-49,300)	1493 (1790; 15-5,343)	3340 (6343; 1 - 49300)

Figure 2. Metadata by top 100 #PTSD TikTok video usefulness group.



Communication Effectiveness, Usefulness Group, and HCP Status

Results are presented in Table 3. There were no group differences across any measures of communication effectiveness, including PEMAT-AV total score (P=.06), PEMAT-AV understandability (P=.12), or the presence of PEMAT-AV actionability points (P=.14). A model with self-identified HCP status and usefulness classification as predictors of communication effectiveness found no significant effect for PEMAT-AV total (P=.38). Self-identified HCPs were responsible for posting 32% (4/12) of misleading content, 14% (9/59) of personal experience content, and 67% (19/29) of useful content, a significant difference between HCPs and non-HCPs (χ^2_2=15.19; P<.001).

Table . Communication effectiveness from top 100 #PTSD TikTok videos and creator status by usefulness group.

	Misleading (n=12)	Personal experience (n=59)	Useful (n=29)	*P* value
PEMAT-AV[a] understandability score, n (%)	58.24 (12.98)	52.63 (12.14)	55.13 (13.05)	.12
PEMAT-AV actionability score >0, %	10	12	33	.14
PEMAT-AV total score, n (%)	46.85 (9.54)	42.11 (10.11)	45.59 (8.37)	.38
Self-identified HCPs[b], %	32	14	67	<.001

[a]PEMAT-AV: Patient Education Materials Assessment Tool for Audiovisual Materials.

[b]HCP: health care professional.

XSL•FO
RenderX

Representations of Trauma Type and *DSM-5* Symptoms

Multimedia Appendix 1 shows the frequency of trauma types and sources represented in our sample, as well as the *DSM-5* symptoms that were mentioned. Of note, the types of traumatic event do not add to 100% because not every creator reported a type of traumatic event.

Discussion

Principal Findings

We found that approximately one-third of the most popular TikTok content related to PTSD contained inaccurate information. The degree to which PTSD-related information was accurate was not associated with its understandability, actionability, or user engagement. Self-identified HCPs were more likely than non-HCPs to post useful information; however, HCPs still accounted for 32% of the content classified to be misleading.

Our findings were roughly consistent with previous studies of mental health–related TikTok content accuracy [10,25]. Studies of other medical conditions reveal variable quality depending on diagnosis: for example, information related to diabetes was generally determined to be of acceptable quality, although lacking in comprehensiveness to meet patient needs [26], and information related to mitral valve regurgitation was generally determined to be factually accurate but lacking in clarity of explicit actionable steps [27]. Of note, TikTok videos are quite short compared to other video-based platforms like YouTube, and it is possible that pressure to fit information in a smaller space leaves less room for misinformation than has been reported to occur there [28]. The content reviewed for this study generally appeared to be informally generated and created by nonexperts or HCPs, so although there was a lack of "actionable steps," this may reflect reluctance to offer advice, which may be considered a positive finding. In addition, the dangers posed by inaccurate information spread may vary according to the characteristics of the condition in question, such as the observability of pathognomonic signs, impact on insight, and implications of meeting diagnostic criteria (eg, prognosis).

In the case of PTSD, this could take the following forms. The misattribution of symptoms with respect to traumatic origin (a problem with specificity) could contribute to both underidentification with trauma disorders (ie, attributing PTSD symptoms to some other cause, such as ADHD), or overidentification with trauma disorders (ie, misattributing nonspecific symptoms to the experience of an adverse event). Previous work shows that group identification (in vs out) is a primary theme of PTSD-specific discussion on internet forums and may serve to reproduce team-based military environments or facilitate "safe" social engagement despite characteristic feelings of detachment or estrangement from other people in general (Criterion D) [11].

Research has shown that aspects of online behavior can become incorporated into users' identities [29]. Once a user is engaged with diagnosis-related content, inaccurate information about prognosis (ie, the common negative belief that people with PTSD are "permanently broken") could endanger one's sense of control, hope for the future, and engagement with treatment. In our view, poor expectations for prognosis are likely the biggest risk from inaccurate information spread about PTSD.

A nonjudgmental approach to practicing clinical psychology in the modern era must acknowledge the very real risks of health misinformation, but also consider the benefits of social media use in providing free, health-related information, and the development of communities around shared symptoms and diagnoses. One such outcome is a sense of support and validation gained through discussions of shared experiences. This could look like a series of reciprocal social interactions (ie, discussions in the comments section under posts), or the building of a community around a particular experience, trend, or content creator. Such community building may function to serve the additional themes of PTSD-specific online discussion identified by Stana et al [11]: overcoming stigma, providing and receiving informational support, and embracing conflict through humor.

For patient populations particularly susceptible to barriers to care related to stigma, changes in diagnostic accuracy over time, and the tendency for symptoms to interrupt engagement with health care information presented in other ways, the delivery of mental health–related information via TikTok is an invaluable resource. For example, a woman in her 60s with a history of learning problems (a problem historically underdiagnosed in girls) and trauma exposure may have never, before accessing TikTok, accessed the information necessary to understand that her day-to-day difficulties may relate to either of these risk factors. If HCPs are acquainted with the digital landscapes inhabited by their patients, these internet resources can be leveraged to extend providers' impact between therapy sessions by supporting patients' critical consumption of this information and even encouraging the discussion of internet content in session.

It is important to consider that TikTok's endless feed represents a massive publicly available source of human behavioral data that may not fit neatly into established categorical systems. It is possible that nonspecific symptoms "overly attributed" to a diagnostic category could reflect a spectrum of patient experiences not captured with longstanding frameworks. The phrase "complex trauma" was specifically mentioned in about one-fourth of our sample. Complex trauma refers to prolonged and repeated exposure to trauma (as opposed to that which occurs via discrete events), for example, exposure to domestic violence, child abuse, human trafficking, and prisoner of war or refugee status [30]. Early research definitions of complex trauma stipulated that the victim is held in a state of physical or emotional captivity by a perpetrator [31].

Complex trauma is not well-captured by current PTSD diagnostic criteria, and the degree to which it may confer a proposed PTSD phenotype known as complex PTSD is the subject of ongoing debate [32,33]. Proponents of a distinct complex PTSD phenotype maintain that diffuse effects of prolonged, repeated trauma on affect regulation, behavioral dysregulation, self-concept, and development (biological and relational) are inefficiently captured and treated by an approach that assigns multiple highly comorbid diagnoses, such as mood

XSL•FO
RenderX

and attachment disorders, ADHD, eating disorders, and sleep disorders. Future work assessing trauma-related content on social media should specifically investigate discussion of complex trauma.

Our study had several strengths. To the best of our ability, we mitigated the effect of our own demographic characteristics on algorithmically driven results of a strategic query by creating a new account and not engaging with features known to impact video recommendations. Also, all human-rated variables were determined via agreement between two independent raters familiar with clinical research related to PTSD, with a third rater resolving rare disagreements between them.

However, our findings should be interpreted in the context of known limitations. Our search approach was strategically intended to capture popular PTSD-related content, but it is not representative of the typical user experience. TikTok uses a proprietary algorithm that is not available to the public but is widely believed to heavily weigh greater user engagement, based on both user experience as well as a leaked internal document confirmed to be a training document for nontechnical employees [34]. According to TikTok's website, the application of its algorithm also differs between users, with the relative weight of indicators of interest, such as video completion or relative geographic proximity of creator and viewer, varying based on assumed importance to the individual user. Second, information about how TikTok's algorithms deliver content through search is not publicly available, and therefore, we cannot know for sure whether these videos represent the most engaged-with.

Conclusions

Approximately one-third of popular PTSD-related TikTok content contained inaccurate information. Accuracy was not linked to content understandability, actionability, or engagement. While self-identified HCPs were more likely to share useful information than nonproviders, they still produced nearly one-third of the misleading content. The democratization of mental health information provides opportunities for individuals who would not otherwise access high-quality care to learn about the experiences and trajectories of others, including free, crowd-sourced tips for managing symptoms. In contrast to other sources of online health information, including other social media platforms, TikTok's primary method of media delivery through an algorithmically determined, continuously adaptive stream of recommended content has the potential to inundate individuals with nonspecific mental health information without users having explicitly searched for it, as might otherwise happen on other sites like YouTube or Facebook. This may, in turn, impact patients' interpretations of their symptoms and ultimately their clinical presentation. Future studies should assess how trends within social media content related to mental health conditions may impact patients' engagement with assessment and treatment within the American health care system.

Data Availability

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

Conflicts of Interest

None declared.

Multimedia Appendix 1
Percentages of trauma and posttraumatic stress disorder symptoms described in the top 100 #PTSD TikTok videos.
[PNG File, 118 KB - ojphi_v17i1e71209_app1.png]

References

1. Woolard A, Paciente R, Munro E, et al. #TraumaTok-TikTok videos relating to trauma: content analysis. JMIR Form Res 2024 Nov 7;8:e49761. [doi: 10.2196/49761] [Medline: 39509697]
2. Milton A, Ajmani L, DeVito MA, Chancellor S. "I see me here": mental health content, community, and algorithmic curation on tiktok. 2023 Apr 19 Presented at: CHI '23; Apr 23-28, 2023; Hamburg Germany. [doi: 10.1145/3544548.3581489]
3. Papacharissi Z. A Networked Self: Identity, Community, and Culture on Social Network Sites: Routledge; 2011:304-318 URL: https://www.researchgate.net/profile/Mark-Andrejevic/publication/49514533_Social_network_exploitation/links/630563051ddd44702104cf6a/Social-network-exploitation.pdf [accessed 2025-08-22]
4. Bhandari A, Bimo S. TikTok and the algorithmized self: a new model of online interaction. Presented at: AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers Virtual Event: AoIR; Oct 27-31, 2020 URL: http://spir.aoir.org [accessed 2025-08-25]
5. Trauma response. Google Trends. URL: https://trends.google.com/trends/explore?q=trauma%20response&date=now%201-d&geo=US&hl=en [accessed 2023-08-24]
6. Palus S. Why TikTok is so obsessed with labeling everything a trauma response. Slate Magazine. 2021 Oct 6. URL: https://slate.com/technology/2021/10/tiktok-trauma-response-why.html [accessed 2025-08-23]
7. Stop your depression by being happy. we know this. TikTok. URL: https://www.tiktok.com/t/ZT6b1nf3k/ [accessed 2025-08-25]

XSL•FO
RenderX

8. Giedinghagen A. The tic in TikTok and (where) all systems go: mass social media induced illness and Munchausen's by internet as explanatory models for social media associated abnormal illness behavior. Clin Child Psychol Psychiatry 2023 Jan;28(1):270-278. [doi: 10.1177/13591045221098522] [Medline: 35473358]
9. Zea Vera A, Bruce A, Garris J, et al. The phenomenology of tics and tic-like behavior in TikTok. Pediatr Neurol 2022 May;130:14-20. [doi: 10.1016/j.pediatrneurol.2022.02.003] [Medline: 35303587]
10. Yeung A, Ng E, Abi-Jaoude E. TikTok and attention-deficit/hyperactivity disorder: a cross-sectional study of social media content quality. Can J Psychiatry 2022 Dec;67(12):899-906. [doi: 10.1177/07067437221082854] [Medline: 35196157]
11. Stana A, Flynn MA, Almeida E. Battling the stigma: combat veterans' use of social support in an online PTSD forum. International Journal of Men's Health 2017;16(1). [doi: 10.3149/jmh.1601.20]
12. Zheng DX, Ning AY, Levoska MA, Xiang L, Wong C, Scott JF. Acne and social media: a cross-sectional study of content quality on TikTok. Pediatr Dermatol 2021 Jan;38(1):336-338. [doi: 10.1111/pde.14471] [Medline: 33247602]
13. Basch CH, Meleo-Erwin Z, Fera J, Jaime C, Basch CE. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. Human Vaccines & Immunotherapeutics 2021 Aug 3;17(8):2373-2377. [doi: 10.1080/21645515.2021.1894896]
14. Russell AM, Davis RE, Ortega JM, Colditz JB, Primack B, Barry AE. #Alcohol: portrayals of alcohol in top videos on TikTok. J Stud Alcohol Drugs 2021 Sep;82(5):615-622. [doi: 10.15288/jsad.2021.82.615] [Medline: 34546908]
15. Leong AY, Sanghera R, Jhajj J, Desai N, Jammu BS, Makowsky MJ. Is YouTube useful as a source of health information for adults with type 2 diabetes? A South Asian perspective. Can J Diabetes 2018 Aug;42(4):395-403. [doi: 10.1016/j.jcjd.2017.10.056] [Medline: 29282200]
16. How to use hashtags on TikTok. #Hashtag Expert. 2023 Jul 16. URL: https://www.hashtag.expert/growth/tiktok-hashtags [accessed 2025-08-23]
17. TikTokPy. Python Package Index. URL: https://pypi.org/project/tiktokapipy/ [accessed 2025-08-23]
18. TikTok scraper. DEV Community. URL: https://dev.to/id1/tiktok-scraper-25a9 [accessed 2025-08-23]
19. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014 Sep;96(3):395-403. [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
20. Sood A, Sarangi S, Pandey A, Murugiah K. YouTube as a source of information on kidney stone disease. Urology 2011 Mar;77(3):558-562. [doi: 10.1016/j.urology.2010.07.536] [Medline: 21131027]
21. Kumar N, Pandey A, Venkatraman A, Garg N. Are video sharing web sites a useful source of information on hypertension? J Am Soc Hypertens 2014 Jul;8(7):481-490. [doi: 10.1016/j.jash.2014.05.001] [Medline: 25064770]
22. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association; 2013. [doi: 10.1176/appi.books.9780890425596]
23. Blom G. Transformations of the binomial, negative binomial, Poisson and $\chi 2$ distributions. Biometrika 1954 Dec;41(3/4):302. [doi: 10.2307/2332711]
24. Mishra S, Diesner J. Detecting the correlation between sentiment and user-level as well as text-level meta-data from benchmark corpora. Presented at: HT '18; Jun 9-12, 2018; Baltimore, MD, USA. [doi: 10.1145/3209542.3209562]
25. Lookingbill V, Mohammadi E, Cai Y. Assessment of accuracy, user engagement, and themes of eating disorder content in social media short videos. JAMA Netw Open 2023 Apr 3;6(4):e238897. [doi: 10.1001/jamanetworkopen.2023.8897] [Medline: 37074713]
26. Kong W, Song S, Zhao YC, Zhu Q, Sha L. TikTok as a health information source: assessment of the quality of information in diabetes-related videos. J Med Internet Res 2021 Sep 1;23(9):e30409. [doi: 10.2196/30409] [Medline: 34468327]
27. Cui N, Lu Y, Cao Y, Chen X, Fu S, Su Q. Quality assessment of TikTok as a source of information about mitral valve regurgitation in China: cross-sectional study. J Med Internet Res 2024 Aug 20;26:e55403. [doi: 10.2196/55403] [Medline: 39163110]
28. Morra S, Collà Ruvolo C, Napolitano L, et al. YouTubeTM as a source of information on bladder pain syndrome: a contemporary analysis. Neurourol Urodyn 2022 Jan;41(1):237-245. [doi: 10.1002/nau.24802] [Medline: 34559920]
29. Kirmayer LJ, Raikhel E, Rahimi S. Cultures of the internet: identity, community and mental health. Transcult Psychiatry 2013 Apr;50(2):165-191. [doi: 10.1177/1363461513490626] [Medline: 23740931]
30. Courtois CA. Complex trauma, complex reactions: assessment and treatment. Psychotherapy: Theory, Research, Practice, Training 2004;41(4):412-425. [doi: 10.1037/0033-3204.41.4.412]
31. Herman JL. Complex PTSD: a syndrome in survivors of prolonged and repeated trauma. J Trauma Stress 1992 Jul;5(3):377-391 [FREE Full text] [doi: 10.1002/jts.2490050305]
32. Cook A, Spinazzola J, Ford J, et al. Complex trauma. Psychiatr Ann 2005;35(5):390-398. [doi: 10.3928/00485713-20050501-05]
33. Resick PA, Bovin MJ, Calloway AL, et al. A critical evaluation of the complex PTSD literature: implications for DSM-5. J Trauma Stress 2012 Jun;25(3):241-251. [doi: 10.1002/jts.21699] [Medline: 22729974]
34. Smith B. How TikTok reads your mind. The New York Times. 2021 Dec 6. URL: https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html [accessed 2025-08-25]

XSL•FO
RenderX

Abbreviations

ADHD: attention-deficit hyperactivity disorder
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
FYP: For Your Page
HCP: health care professional
PEMAT-AV: Patient Education Materials Assessment Tool for Audiovisual Materials
PTSD: posttraumatic stress disorder

Edited by E Mensah; submitted 15.01.25; peer-reviewed by J Mandzufas, N Maamor, R Nordgren, R Kumar, S Morra; revised version received 10.04.25; accepted 14.06.25; published 10.09.25.

Please cite as:
Rohl B, Jones LC, Nattis R, Claar RD, Velez X, Gabrielli J, Williamson J, Porges E
Posttraumatic Stress Disorder Content on TikTok: Cross-Sectional Analysis of Popular #PTSD Posts
Online J Public Health Inform 2025;17:e71209
URL: https://ojphi.jmir.org/2025/1/e71209
doi:10.2196/71209

© Brittany Rohl, Laura Carolyn Jones, Rachel Nattis, Robert Dale Claar, Xavier Velez, Joy Gabrielli, John Williamson, Eric Porges. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 10.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Expansion and Assessment of a Web-Based 24-Hour Dietary Recall Tool, Foodbook24, for Use Among Diverse Populations Living in Ireland: Comparative Analysis

Grace Bennett[1], PhD; Shuhua Yang[1], PhD; Laura A Bardon[2], PhD; Claire M Timon[3], PhD; Eileen R Gibney[1], PhD

[1]UCD Institute of Food and Health, University College Dublin, Dublin, Ireland

[2]Quadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom

[3]School of Population Health, Royal College of Surgeons, Dublin, Ireland

Corresponding Author:
Eileen R Gibney, PhD
UCD Institute of Food and Health
University College Dublin
Science Centre South
Belfield
Dublin, 4
Ireland
Phone: 353 17161147
Email: eileen.gibney@ucd.ie

Abstract

Background: Currently, the methods used to collect dietary intake data in Ireland are inflexible to the needs of certain populations, who are poorly represented in nutrition and health data as a result. As the Irish population is becoming increasingly diverse, there is an urgent need to understand the habitual food intake and diet quality of multiple population subgroups, including different nationalities and ethnic minorities, in Ireland. Foodbook24 is an existing web-based 24-hour dietary recall tool, which has previously been validated for use within the general Irish adult population. Because of its design, Foodbook24 can facilitate the improved inclusion of dietary intake assessment in Ireland.

Objective: We aimed to examine the suitability of expanding the Foodbook24 tool, improving the reliability and accuracy of dietary intake data collected among prominent nationalities in Ireland.

Methods: This study consisted of three distinct parts: (1) expansion of Foodbook24, (2) testing its usability (ie, acceptability study), and (3) examining the accuracy (ie, comparison study) of the updated Foodbook24 tool. To expand Foodbook24, national survey data from Brazil and Poland were reviewed and commonly consumed food items were added to the food list. All foods were translated into Polish and Portuguese. The acceptability study used a qualitative approach whereby participants provided a visual record of their habitual diet. The comparison study consisted of one 24-hour dietary recall using Foodbook24 and one interviewer-led recall completed on the same day, repeated again 2 weeks later. Comparison study data were analyzed using Spearman rank correlations, Mann-Whitney *U* tests, and κ coefficients.

Results: The expansion of the Foodbook24 food list resulted in 546 additional foods. The acceptability study reported that 86.5% (302/349) of foods listed by participants were available in the updated food list. From the comparison study, strong and positive correlations across 8 food groups (44% of a total of 18 food groups) and 15 nutrients (58% of a total of 26 nutrients) were identified (r=0.70-0.99). Only intakes of *potatoes and potato dishes* and *nuts, herbs, and seeds* significantly differed across methods of assessment, where correlations across these food groups were low (r=0.56 and r=0.47, respectively). The incidence of food omissions varied across samples, with Brazilian participants omitting a higher percentage of foods in self-administered recalls than other samples (6/25, 24% among the Brazilian vs 5/38, 13% among the Irish cohort).

Conclusions: The updated food list is representative of most foods consumed by Brazilian, Irish, and Polish adults in Ireland. Dietary intake data reported in Foodbook24 are not largely different from food groups and nutrient intakes reported via traditional methods. This study has demonstrated that Foodbook24 may be appropriate for use in future research investigating the dietary intakes of Brazilian, Irish, and Polish groups in Ireland.

(Online J Public Health Inform 2025;17:e52380) doi:10.2196/52380

XSL•FO
RenderX

KEYWORDS

dietary assessment; ethnic diets; dietary records; web-based tools; diverse intake

Introduction

Background

Following a call from the European Food Safety Authority seeking "initiatives to further develop web-based tools in the area of dietary surveys," the development of technology-based dietary assessment tools has greatly increased [1]. Traditional paper-based methods of dietary assessment are inflexible, costly, and labor intensive for both participants and researchers [2,3]. Many recently developed web-based tools offer improved accuracy and provide researchers with additional flexibility to capture data from those living abroad, who have reduced mobility or limited access to transport, or who are not proficient in the same language as the researcher [4-6]. Although digital advances have improved how dietary intake is assessed, systematic errors such as recall bias, lack of diversity within food lists, lack of flexibility for different languages, and inaccurate food portion size estimates remain [7-9]. Similar to many high-income countries across Europe and worldwide, the demographic landscape of Ireland is changing rapidly [10,11]. Not only is the Irish population aging, but also it is becoming more culturally diverse, with the number of different ethnic groups living in Ireland growing by approximately 20% annually [10,12]. As a result, the country has been enriched with an increased array of languages, religions, and cultural and dietary practices. Despite this, it is widely accepted that ethnic minority groups remain underrepresented in social and nutritional science research [13,14]. Ensuring diversity in ongoing data collection is important to understand current dietary patterns of different population subgroups and identifying if targeted interventions are required to limit diet-associated health risks.

Diversity in Nutrition Data

Food-based dietary guidelines primarily rely on data from national food consumption surveys, but methods of dietary assessment used in these surveys are often not adapted to include diverse languages and foods consumed by different nationalities. In essence, national food consumption surveys tend to underrepresent specific population subgroups [15]. Some national food consumption surveys look to cater to diverse population groups by offering alternative languages for specific samples: in the United States, the National Health and Nutrition Examination Survey food frequency questionnaire is also translated into Spanish for Hispanic participants, and in Switzerland, the menuCH 24-hour dietary recall is available in German, French, or Italian; however there is no such dietary assessment tool available for the diverse population in Ireland [16,17]. In Ireland, adult national food consumption surveys are conducted every 10 years and while data on ethnicity are collected, very poor diversity is achieved; therefore, little is known about the dietary intakes of the ethnic minority groups in Ireland and how they compare to each other and to the healthy eating guidelines.

Foodbook24, a web-based 24-hour dietary recall tool, was originally developed to accurately assess dietary intakes of the Irish adult population living in Ireland [18,19]. As with many web-based dietary assessment tools, participants select foods consumed from a prepopulated food list; nutrient compositions are based on the United Kingdom "Composition of Food Integrated Database" (CoFID), and food images are used to estimate portion sizes [19]. Foodbook24 has been previously validated against 4-day food diaries, interviewer-led 24-hour dietary recalls, and biomarkers of dietary intake (ie, using plasma and urine samples) [19,20]. Foodbook24 was designed for the general Irish population: English speakers who follow a typical "Western" diet, and the original food list was based on foods reported by Irish adults as part of the National Adult and Nutrition Survey; therefore, its use in population subgroups such as specific nationalities living in Ireland who consume a diverse range of food items is limited.

Objectives

To further enhance the capabilities of the existing tool and ensure it is appropriate for use among different nationalities living in Ireland, Foodbook24 was further developed to include more languages, including Brazilian Portuguese and Polish, and food items commonly consumed by Polish and Brazilian adults. This study aimed to examine the suitability of expanding the Foodbook24 tool for use among diverse nationalities in Ireland and assess the accuracy of the dietary intake data collected via Foodbook24 by comparing it to traditional assessment methods, ensuring dietary intake of the ethnic minority groups can be investigated using appropriate assessment tools.

Methods

Selection of Diverse Population Groups

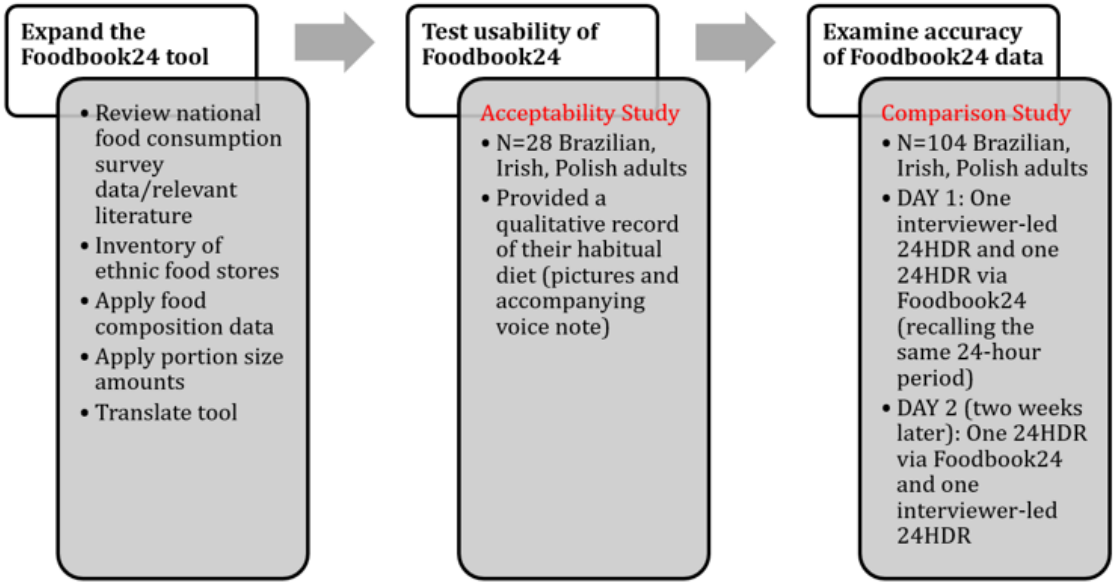
The samples selected within this study were chosen to ensure representation of the current population of Ireland. Irish adults were included as a reference sample, as the tool was previously shown to be comparable within this sample [1]. Renewed comparison among this group would ensure the food list was still reflective of Irish dietary habits today. Polish and Brazilian samples were chosen on the basis of their prevalence in Ireland, differing native languages, and unique dietary traditions [12]. According to the latest available census from 2016, the largest population group in Ireland apart from Irish and Irish Traveler was the Polish population, who equated to 2.6% (122,515/4,761,865) of the total Irish population [12]. Other than British and Eastern European residents, who follow similar dietary patterns to Irish and Polish groups, Brazilians were the next most common non-Irish nationality living in Ireland (approximately 14,000 nationals) [12]. In addition to a large number of people from these countries living in Ireland, these samples are distinctly and culturally different from Ireland and from each other in terms of their native tongue and habitual dietary intake.

This study consisted of three distinct parts: (1) expansion of Foodbook24, (2) testing the usability of the Foodbook24 tool (ie, acceptability study), and (3) examining the accuracy of the

XSL•FO
RenderX

updated Foodbook24 tool (ie, comparison study). Each of these sections are described in detail in the subsequent sections, individually. Figure 1 provides a summary of each part of the Foodbook24 expansion and testing as well as key activities within each phase.

Figure 1. Summary of the 3 phases of Foodbook24 expansion. 24HDR: 24-hour dietary recall.



Expansion of Foodbook24

Expansion of Foodbook24 was conducted in a multistage process to ensure it could accurately cater for specific population groups (ie, diverse nationalities, vegans, and vegetarians). The first stage of expanding Foodbook24 was to identify and add food items and beverages frequently consumed by Brazilian and Polish adults according to national food consumption surveys and relevant literature from each country [21-29]. The food list in Foodbook24 was also updated to ensure that it fully reflects the current food consumption habits of Irish adults. The original food list in Foodbook24 was based on foods reported in National Adult and Nutrition Survey and had not been updated for the Irish adult population since 2017 [18,19]. On the basis of recent literature and findings from the Foodbook24 acceptability study, there has been a rise in the consumption of certain food products, as well as vegetarianism and veganism, in Ireland. These new-to-market foods were also added to the Foodbook24 food list to account for their increased consumption in Ireland. Next, nutrient composition data were applied to the newly added foods. The Foodbook24 food list primarily consists of individual food items, rather than composite meals, allowing participants to account for cultural-specific recipes and meal ingredients. As the foods listed in Foodbook24 are available on the Irish market, CoFID nutrient compositions were applied to foods where possible [30]. National nutrient composition databases from Brazil and Poland were used for foods consumed by these populations where nutritional information did not exist in CoFID (ie, foods which are likely more culturally specific and less widely available in Ireland) [31,32]. For a limited number of non-Polish or Brazilian food items added to the Foodbook24 food list, predominantly those considered new to the food market (eg, many milk alternatives such as oat or almond), no appropriate composition could be found within CoFID; therefore, average food label information from 5 branded sources were used. Portion size estimation was then applied to new foods; portion size estimates were sought from national food consumption surveys, and if unavailable, portion size estimates of the closest alternative food or drink item from the Ministry of Agriculture, Fisheries and Food's "Food Portion Sizes" book were used [33]. From national food consumption surveys, the mean reported intake was defined as the medium portion size. One SD was then added to this to determine the medium-large portion size, 2 SDs were added for the large portion size, and 3 SDs were added for the extra-large portion. To calculate smaller portion sizes, increasing SDs were subtracted from the mean reported intake (ie, mean–1 SD, mean–2 SD, and mean–3 SD for the medium-small, small, and extra small portion sizes, respectively). For some foods that were consumed less frequently, the SD was larger than the mean reported intake; therefore, the smaller and larger portion sizes were calculated as the mean +25% to –25%. Existing portion size images were used where possible for the newly added food and drink items. Where there were no appropriate pre-existing images, new food portion size images were taken following a standardized procedure. Finally, all aspects of the Foodbook24 tool were translated and tested by native Brazilian Portuguese and Polish speakers.

Acceptability Study Design

To test the completeness of the expanded food list, an acceptability study was conducted. The main aim of the

XSL•FO
RenderX

acceptability was to understand habitual diets of diverse nationalities living in Ireland and assess the ease at which these diets would be captured using Foodbook24. Participants (n=28) were asked to complete a collage of their habitual diet using the web-based application Pinterest (Pinterest, Inc) and record an accompanying voice note providing further details of the food items and beverages reported within the Pinterest board such as cooking method, condiments added, and whether they were homemade or ready-made food items. This qualitative method of describing habitual diet, commonly termed as *photovoice*, has been previously used in the form of focus groups to identify food choice and practices of underrepresented communities [34,35]. As this study was conducted during the COVID-19 lockdown, it was not possible to conduct focus groups; therefore, the approach using Pinterest was preferred. For the analysis of the acceptability data, the food content of each Pinterest board was extracted, along with the accompanying voice notes, which were transcribed verbatim for clarity and to support the content. Two researchers (GB and SY) extracted data collected from the acceptability study and formed a list of the final food items reported by participants. All authors reviewed this list and a consensus was then reached as to what foods to add to Foodbook24 on the basis of (1) the frequency each food item was reported and (2) the similarity of each food to the ones existing in Foodbook24 (ie, foods deemed similar were not added as separate food items but were "tagged" to existing foods).

Comparison Study Design

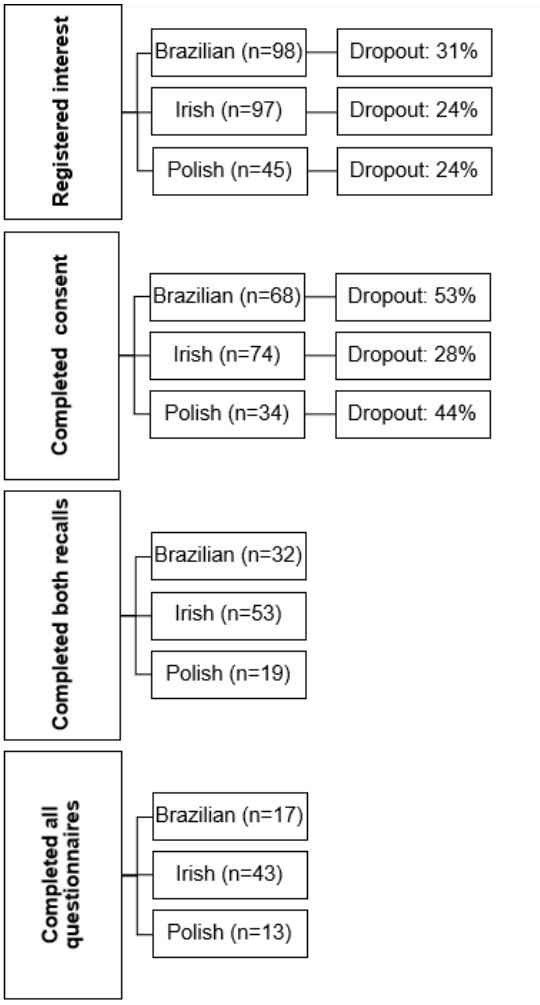
The same protocol that was originally used to assess the use of Foodbook24 among the Irish adult population was applied to this study (Figure S1 in Multimedia Appendix 1) [20]. At baseline, participants (n=104) completed one self-administered 24-hour dietary recall using Foodbook24 and one interviewer-led 24-hour dietary recall with a trained researcher on the same day. The order in which recalls were completed was randomized, with 75% (78/104) of the participants completing the Foodbook24 recall first. Two weeks later, the process was repeated in the opposite order to the first dietary recall. Interviewer-led recalls were conducted by 2 researchers with a strong background in nutrition or dietetics using the Zoom platform (Zoom Communications, Inc) because of COVID-19 pandemic restrictions (GB and SY). Researchers followed a predefined protocol to ensure consistency across interviewer-led recalls, which was developed in accordance with the automated multiple pass method by the US Department of Agriculture [36]. Through previous research activities, researchers conducting the interviewer-led recalls had a strong understanding of dietary cultures of each sample group. All interviews were conducted in English. During the interviewer-led recalls, Foodbook24 portion size images within a food atlas were shown to participants using the "share screen" function on Zoom to estimate portion size using the same method as the self-administered method. If a food item mentioned by participants during the interviewer-led recall was not available in the food list, portion size images of a similar food item were used instead. For example, the portion size image for "white rolls" was shown to participants when the food item *pan de queso* (ie, cheese bread) was mentioned. Participant intake data collected through interviewer-led recalls were recorded on paper. The data were subsequently entered by the same researcher who facilitated the interviewer-led recall into Foodbook24 so that the 2 methods could be compared; by re-entering interviewer-led recalls into Foodbook24, no differences were found in portion size estimates or food composition data. This was important to ensure that the differences in reported intakes could be explained by different assessment approaches rather than other aspects of the 24-hour dietary recalls; the previous Foodbook24 comparison study also used this method [20]. Participants also completed demographic and tool evaluation questionnaires as part of this study.

Participant Recruitment

Recruitment for both acceptability and comparison studies took place between January 2021 and May 2022. During recruitment, authors hoped to achieve an equal distribution of Brazilian, Irish, and Polish participants across representative age and sex categories. In total, 132 participants were recruited, including 37 (28%) Brazilian, 74 (56.1%) Irish, and 21 (15.9%) Polish. Social media, in-person, and web-based avenues were used to recruit target groups. To attract attention from within the community at University College Dublin (UCD), study posters were displayed around the campus and recruitment events hosted. Other relevant in-person events across Dublin were also attended, such as a Brazilian craft fair and an exhibition for older adults in Dublin. Study information in the form of flyers was distributed outside food shops and restaurants owned by Polish and Brazilian populations. Information posters were also hung on noticeboards of churches, where services were held in Polish, and community centers. Study posters and flyers were printed with an active QR code so that participants could sign up and begin the study directly. Study information was also shared on the web on the UCD Institute of Food and Health and Irish Volunteer websites. Study-specific Facebook (Meta Platforms, Inc) and Instagram (Meta Platforms, Inc) accounts were created to attract attention among target groups, and advertisements were posted on Facebook. Participants were deemed eligible once they identified as a Brazilian, Irish, or Polish adult (aged ≥18 years), lived in Ireland, spoke basic English (ie, able to communicate with researchers during interviewer-led recalls through English), and did not hold a formal degree in nutrition or dietetics. An overview of participant recruitment and retention numbers for the comparison study is available in Figure 2.

XSL•FO
RenderX

Figure 2. Participant interest and retention at each stage of the comparison study.



Statistical Analysis

Descriptive analysis was performed on qualitative acceptability study data to understand the food type and frequency of reporting among target groups of interest. Foods recorded by participants in the acceptability study were compared to foods listed in Foodbook24. Mean daily energy, macro- and micronutrient, and food group intakes from the comparison study were compared across the 2 methods (ie, self-administered and interviewer-led recalls). The relationship between both methods was compared using Spearman rank correlations and percentage differences [(self-administered intake − interviewer-led intake) / (self-administered) × 100]. Data were not normally distributed, (ie, according to the Kolmogorov-Smirnov test; $P<.001$); thus, nonparametric tests were used for analysis. Mann-Whitney *U* tests were performed to identify significant differences between methods (ie, $P\leq.05$). Statistical significance was assessed using unadjusted and Bonferroni-corrected *P* values [35]. Bland Altman analysis was conducted for key nutrients and food groups to estimate the interval of agreement (95% CI), where the self-administered nutrient value was subtracted from the interviewer-led nutrient value. Relative agreement of the 2 recall methods was assessed using cross-classification analysis. Agreement across methods was also assessed by examining the frequency of food matches, omissions, and intrusions rates for the total population and individual samples [20,36,37]. Exact matches were items that were identical in both recalls, and total matches included exact matches as well as those that were a similar match (eg, low-fat vs full-fat milk). Omissions were defined as items recorded in the interviewer-led recall but not the self-administered recall, and intrusions were those reported in the self-administered recall but not the interviewer-led recall [20]. Misreporters are removed from nutritional data to provide mean intakes of a sample who provided a likely accurate report of what they ate during the 24-hour period. To identify intake underreporters, estimated energy requirements were calculated by multiplying the basal metabolic rate (ie, based on the Henry equation) by 1.1 [38]. Participants whose mean intakes were below their estimated energy requirements for both interviewer-led and self-administered 24-hour dietary recalls were deemed to be underreporters. Only data for adequate reporters are presented here; however, results for the total population including underreporters and individual cohorts are available in Tables S1-S15 in Multimedia Appendix 1. Overreporters were also considered in this research by applying energy intake cutoffs, but none were identified in this sample. All data analyses were performed using SPSS Statistics (version 27; IBM Corp), and statistical significance was set at $P<.05$.

XSL•FO
RenderX

Ethical Considerations

Ethics approval for both the acceptability and comparison studies was provided by the Office of Research Ethics in UCD (LS-20-48-Gibney). All participant information leaflets and consent forms were provided to participants in English and their native language of Brazilian or Polish. Within the participant information leaflets, participants were informed what data would be collected and why, how these data would be handled and used (ie, pseudonymized immediately once collected), and how they could withdraw themselves and their data from the study if desired. Participants could only participate in this research once the consent form was complete. All participant data was fully de-identified before analysis where each participant was given an unique identifier code. There were no financial incentives for participants to take part in this research. Participants were offered personalized dietary feedback after their completion of the study.

Results

Foodbook24 Expansion

During this phase of Foodbook24 development, a total of 546 new foods were added to the Foodbook24 food list; of these, 420 (76.9%) were ethnic foods and 126 (23.1%) were additional Irish foods. The 24-hour dietary recalls in Foodbook24 can now be accurately completed in English, Polish, and Portuguese where participants can select their language of preference. All data outputs available to researchers are presented in English.

Acceptability Study

In total, 28 participants completed the acceptability study; of these, 21 (75%) were Irish, 5 (18%) were Brazilian, and 2 (7%) were Polish. The mean age of the total sample was 37.71 (SD 16.48) years, with participants from Polish and Brazilian samples aged between 27 and 41 years. In total, 349 items were recorded by participants in the acceptability study, including 65 (18.6%) from *breads, potatoes, and grains*; 40 (11.5%) from *fats, sugars, and desserts*; 17 (4.9%) from *dairy*; 82 (23.5%) from *protein foods*; 78 (22.3%) from *fruit and vegetables*; and 67 (19.2%) from the *other* group. Of the 349 items mentioned, 302 (86.5%) were available in the Foodbook24 food list or a similar alternative to the reported item was available (eg, *garlic mayonnaise* did not exist in the food list but could be substituted for *mayonnaise, regular fat*). In this scenario, names of items not explicitly included in the food list but compositionally very similar to existing foods were added as *search tags* to these original items within Foodbook24: search tags are linked to food items in Foodbook24 to support participants to find foods quickly; search tags include brand names or common colloquial terms for food items. Frequently reported food and drink items, which were not explicitly named in Foodbook24, were predominately from the *soups, sauces, and miscellaneous* and *cheese* groups (14/44, 32% and 3/10, 30% of reported foods from these groups, respectively, were not included in the food list). From the *soups, sauces, and miscellaneous* group, condiments such as coconut oil and peri peri sauce were added as *search tags* as suitable alternatives were available in the Foodbook24 food list. These foods, which were similar to existing foods in terms of type and composition, were not added to the updated Foodbook24 food list to keep the Foodbook24 tool concise and not to overwhelm or confuse participants when selecting food items. Edam and Gouda were popular cheeses reported by Brazilian and Polish adults; as these items were not specifically mentioned in the food list, they were added as new foods.

Comparison Study

In total, 104 participants were recruited to take part in the comparison study, including 53 (51%) Irish, 32 (31%) Brazilian, and 19 (18%) Polish participants. The mean age of the total sample was 38.83 (SD 14.19) years, 72% (75/104) of which were female participants (Table 1). Findings from the total population are presented in Tables 2-4. Findings for each individual sample (Brazilian, Irish, and Polish) are available in Tables S2-S4 in Multimedia Appendix 1. Comparing demographic characteristics across sample groups, Brazilian and Polish groups were predominately female (28/32, 88%) and 79% (15/19) of Brazilian and Polish adults were aged between 30 and 45 years (18/32, 56% and 12/19, 63% respectively). The Irish sample was more evenly distributed in terms of sex (32/53, 60% female) and age (7/53, 13% aged 30-45 years, ranging from 18 to 72 years). Full demographic data were obtained for 84 participants (30/85, 36% underreporters; 54/85, 64% adequate reporters; and no overreporters were identified), where 74% (63/85) have completed third-level education (minimum of primary bachelor's degree). Of the total sample, 30 (29%) were considered underreporters and so were excluded from results reported in Tables 2-5. Comparing demographic characteristics between adequate and underreporters, underreporters were more likely to be male (underreporters 13/30, 43% vs adequate reporters 13/54, 24%), older (mean age of under reporters43.07, SD 15.21 years vs mean age of adequate reporters=36.54, SD 13.86 years), and have a third-level education (underreporters 24/30, 80% vs adequate reports 39/54, 72%; Table S1 in Multimedia Appendix 1). Underreporters were more likely to have overweight status (mean BMI 26.39 kg/m^2) compared to adequate reporters whose mean weight status was in the normal category (mean BMI 23.41 kg/m^2; Table S1 in Multimedia Appendix 1).

Total sample mean nutrient intakes from the self-administered Foodbook24 method were compared to mean nutrient intakes from the interviewer-led method, excluding underreporters (Table 2). Minimal differences in key nutrient intakes were observed between both methods. Vitamin A was the only nutrient where a statistically significant difference was observed, whereby mean vitamin A intake via the interviewer-led method was significantly higher than the self-administered recall (mean 1050.05, SD 589.57 µg/day vs mean 907.47, SD 702.54 µg/day; P=.03; r=0.61). Moderate to high associations, using Spearman rank correlation analysis, for all nutrients were found between both recall methods with correlations ranging from 0.51 to 0.77 (Table 2).

XSL•FO
RenderX

Table 1. Participant demographics of the total sample from the comparison study.

Participant characteristics	Total sample	Brazilian	Irish	Polish	*P* value
Population, n (%)[a]	104 (100)	32 (21.8)	53 (36.1)	19 (12.3)	—[b]
Age (y), mean (SD)[a]	38.83 (14.2)	33.19 (5.8)	41.89 (18.1)	39.89 (8.1)	.02
Sex, n (%)[a]					.09
Male	29 (27.9)	4 (12.5)	21 (40.0)	4 (21.0)	
Female	75 (72.0)	28 (87.5)	32 (60.0)	15 (79.0)	
BMI (kg/m^2)[c]	23.51	22.40	23.67	24.88	.53
Education, n (%)[c]					.21
Level ≥8[d]	63 (74.0)	16 (73.0)	36 (75)	11 (79.0)	
Other[e]	22 (26.0)	6 (27.0)	12 (25)	3 (21.0)	
Physical activity, n (%)[c]					.06
≤2 times, weekly	35 (42.0)	11 (50.0)	20 (43.5)	4 (29.0)	
≥3 times, weekly	48 (58.0)	11 (50.0)	26 (56.5)	10 (71.0)	

[a]Percentage of participants within the total sample, not those who completed the full demographic questionnaire.

[b]Not applicable.

[c]BMI, education, and physical activity data were provided by 84 participants who completed the full demographic questionnaire.

[d]Primary or postgraduate university degree.

[e]Leaving certificate degree (high school equivalent), diploma, or apprenticeship.

XSL•FO
RenderX

Table 2. Mean energy and nutrient intakes of adequate reporters recorded via interviewer-led and self-administered methods (comparison study).

Nutrients	Interviewer-led intakes (n=74)[a], mean (SD)	Self-administered intakes (n=74)[b], mean (SD)	*P* value[c]	Difference (%)	Correlations (*r*[d])
Energy (kcal/day)	2000.95 (526.66)	2077.61 (724.51)	.71	3.69	0.63
Protein (g/day)	87.56 (36.83)	83.86 (30.61)	.72	–4.40	0.52
Carbohydrate (g/day)	235.05 (76.64)	245.54 (91.23)	.54	4.27	0.64
Sugars (g/day)	96.20 (46.27)	100.30 (64.11)	.86	4.08	0.77
Starch (g/day)	130.53 (45.21)	130.99 (51.64)	.99	0.36	0.58
Dietary fiber (g/day)	18.19 (6.92)	17.84 (6.80)	.83	–1.93	0.62
Total fat (g/day)	82.44 (23.82)	84.17 (27.46)	.81	2.05	0.64
Saturated fat (g/day)	32.00 (12.66)	31.72 (14.63)	.70	−0.86	0.77
Monounsaturated fat (g/day)	29.51 (9.41)	29.39 (11.18)	.89	−0.38	0.68
Polyunsaturated fat (g/day)	6.40 (2.68)	6.88 (3.56)	.77	6.93	0.51
Protein (% energy)	17.74 (6.13)	16.59 (4.67)	.45	−6.96	0.63
Carbohydrate (% energy)	43.80 (7.01)	44.28 (6.87)	.60	1.10	0.68
Total fat (% energy)	37.28 (6.11)	37.14 (7.06)	.90	−0.38	0.75
Vitamin D (µg/day)	7.48 (8.37)	7.95 (9.41)	.72	5.89	0.70
Vitamin E (mg/day)	15.39 (12.25)	14.35 (10.52)	.58	−7.24	0.70
Vitamin B6 (mg/day)	4.91 (10.66)	4.29 (8.53)	.97	−14.43	0.71
Vitamin B12 (µg/day)	19.51 (115.67)	14.69 (58.66)	.62	−32.80	0.59
Folate (mg/day)	350.28 (224.29)	347.15 (240.54)	.58	−0.90	0.79
Vitamin C (mg/day)	208.51 (234.55)	152.23 (192.50)	.13	−36.97	0.63
Calcium (mg/day)	992.25 (382.92)	923.89 (399.07)	.20	−7.40	0.65
Magnesium (mg/day)	332.42 (110.58)	353.96 (164.43)	.76	6.08	0.68
Phosphorus (mg/day)	1353.13 (357.09)	1370.17 (541.70)	.72	1.24	0.68
Iron (mg/day)	19.07 (22.03)	18.40 (19.90)	.83	−3.65	0.63
Iodine (mg/day)	161.23 (94.58)	145.24 (93.79)	.20	−11.01	0.69
Vitamin A (µg/day)	1050.05 (589.57)	907.37 (702.54)	.03[e]	−15.73	0.61
Carotene (µg/day)	3934.75 (3227.69)	3288.09 (3483.47)	.06	−19.67	0.56

[a]Interviewer-led 24-hour dietary recall.

[b]Self-administered 24-hour dietary recall (via Foodbook24).

[c]*P* values represent significant differences between methods (Mann-Whitney *U* tests); significance at <.05 level.

[d]All *P* values for Spearman rank correlations were <.001.

[e]Only significant for unadjusted *P* values following Bonferroni correction.

XSL•FO
RenderX

Table 3. Mean food group intakes of adequate reporters recorded via interviewer-led and self-administered methods (comparison study).

Food group (g/day)	Interviewer-led intakes (n=74)[a], mean (SD)	Self-administered intakes (n=74)[b], mean (SD)	*P* value[c]	Difference (%)	Correlations (r[d])
Beverages[e]	997.14 (894.54)	926.92 (1357.55)	.07	−7.58	0.67
Biscuits, cakes, and buns	46.89 (35.62)	51.63 (48.26)	.96	9.19	0.75
Breads, rolls, and scones	74.94 (43.41)	66.06 (43.78)	.21	−13.45	0.63
Breakfast cereals	103.22 (84.83)	104.09 (85.84)	.99	0.83	0.96
Butter, spreads, and oils	11.16 (6.96)	10.85 (7.69)	.55	−2.87	0.51
Cheese	27.20 (21.76)	20.99 (13.58)	.19	−29.58	0.53
Creams, ice creams, and desserts	65.64 (58.84)	51.45 (47.69)	<.001[f]	−27.59	0.71
Egg and egg dishes	107.64 (85.38)	105.25 (100.42)	<.001[f]	−2.27	0.76
Fish and fish dishes	64.67 (55.92)	73.38 (64.87)	.06	11.87	0.69
Fruit and fruit juices	271.09 (251.96)	266.29 (218.23)	<.001[f]	−1.80	0.86
Grains, rice, pasta, and savories	161.28 (105.43)	174.78 (100.52)	.02[g]	7.72	0.58
Meat and meat products	155.41 (94.98)	163.81 (97.66)	.38	5.13	0.68
Milk and yoghurts	140.26 (108.36)	145.00 (139.50)	.06	3.27	0.66
Nuts, herbs, and seeds	9.57 (7.08)	14.09 (10.74)	<.001[f]	32.07	0.47
Potatoes and potato dishes	98.62 (51.71)	112.30 (66.97)	<.001[f]	12.18	0.56
Soups, sauces, and miscellaneous	96.88 (110.01)	97.31 (110.58)	0.01[g]	0.44	0.74
Sugars, confectionary, and preserves	38.35 (35.15)	39.97 (38.78)	0.02[h]	4.05	0.86
Vegetables and vegetable dishes	200.35 (126.47)	182.61 (147.38)	<.001[h]	−9.71	0.79

[a]Interviewer-led 24-hour dietary recall.

[b]Self-administered 24-hour dietary recall (via Foodbook24).

[c]*P* values represent differences between methods (Mann-Whitney *U* tests).

[d]All *P* values for Spearman rank correlations were <.001.

[e]Water removed from *beverages* food category.

[f]Only significant for unadjusted *P* values following Bonferroni correction.

[g]Significant for adjusted and unadjusted *P* values following Bonferroni correction.

[h]Only significant for unadjusted *P* values following Bonferroni correction.

XSL•FO
RenderX

Table 4. Cross-classification of mean food group intakes of adequate reporters recorded via self-administered and interviewer-led methods (comparison study).

Food group	κ coefficient	Exact[a] (%)	Adjacent[b] (%)	Disagreement[c] (%)	Extreme disagreement[d] (%)
Beverages[e]	0.50	62.16	28.38	6.76	2.70
Biscuits, cakes, and buns	0.48	64.86	27.03	5.41	2.70
Breads, rolls, and scones	0.33	52.70	36.49	6.76	4.05
Breakfast cereals	0.87	91.89	8.11	0	0
Butter, spreads, and oils	0.46	63.51	22.97	9.46	4.05
Cheese	0.45	67.57	20.27	6.76	5.41
Creams, ice creams, and desserts	0.49	79.73	13.51	2.70	4.05
Egg and egg dishes	0.73	90.54	4.05	4.05	1.35
Fish and fish dishes	0.60	85.14	13.51	0	1.35
Fruit and fruit juices	0.71	78.38	18.92	1.35	1.35
Grains, rice, pasta, and savories	0.46	59.46	28.38	9.46	2.70
Meat and meat products	0.34	51.35	36.49	9.46	2.70
Milk and yoghurts	0.33	56.76	25.68	12.16	5.41
Nuts, herbs, and seeds	0.46	68.92	24.32	2.70	4.05
Potatoes and potato dishes	0.56	72.97	16.22	6.76	4.05
Soups, sauces, and miscellaneous	0.38	55.41	36.49	8.11	0
Sugars, confectionary, preserves	0.47	62.16	36.49	1.35	0
Vegetables and vegetable dishes	0.45	58.11	35.14	6.76	0

[a]Percentage of cases cross-classified into the same quartile.

[b]Percentage of cases cross-classified into an adjacent quartile (+1 or −1).

[c]Percentage of cases cross-classified 2 quartiles apart.

[d]Percentage of cases cross-classified 3 quartiles apart.

[e]Water removed from the *beverage* food group.

Table 5. Percentage of matches, omissions, and intrusions observed between interviewer-led and self-administered recalls across samples (comparison study).

	Total match[a], n (%)	Exact match[b], n (%)	Omissions[c], n (%)	Instrusions[d], n (%)
Brazilian (n=25)	17 (69)	13 (52)	6 (23)	2 (8)
Irish (n=38)	31 (81)	25 (67)	4 (12)	3 (7)
Polish (n=11)	8 (74)	6 (56)	2 (18)	1 (8)
Total sample (n=74)	55 (74)	43 (58)	13 (18)	6 (8)

[a]Total matches include items that were identical as well as similar, for example, whole meal bread and granary bread were recorded as similar matches.

[b]Exact matches represent items that were identical only.

[c]Omissions account for items reported in the interviewer-led recall but not the self-administered recall.

[d]Instrusions include items reported in the self-administered recall but not the interviewer-led recall.

Within the total sample, associations among total sugars, fat (as a percentage of energy), saturated fat, and folate were particularly strong. Bland Altman plots display agreement across key nutrients (Figures 3-8). Good agreement was found for protein (percentage of total energy), sugar, and calcium (Figures 3, 6, and 8, respectively) with <5% of participants falling outside the limits of agreement. Approximately 5% of participants fell outside the limits of agreement for carbohydrate (% energy) and total fat (% energy); however, correlations for these nutrients were moderate (r=0.6) and high (r=0.90), respectively (Figures 4 and 5; Table 2).

XSL•FO
RenderX

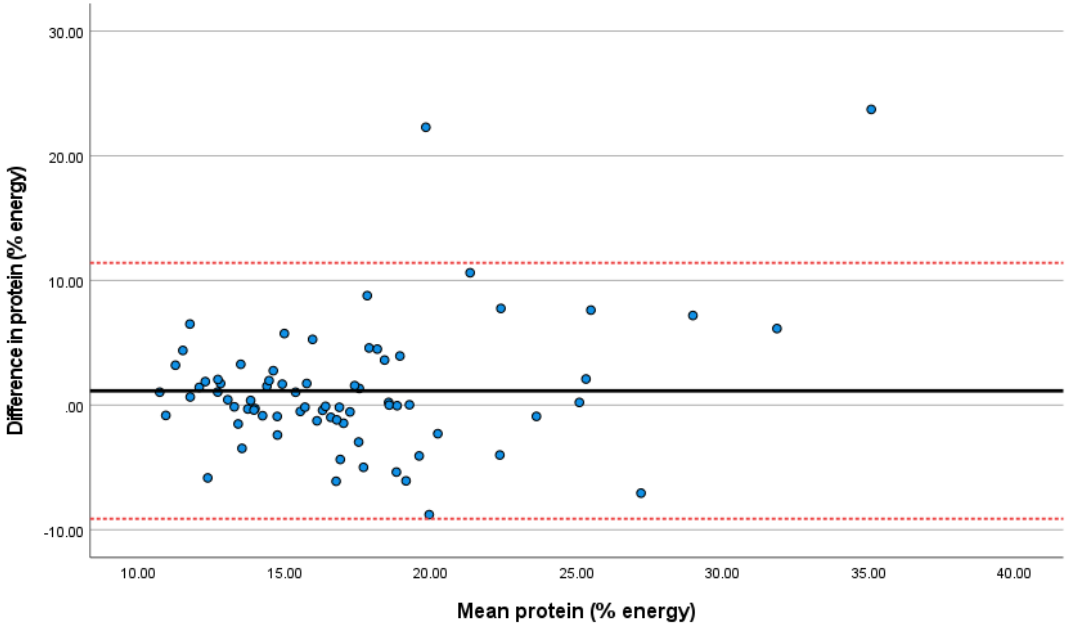
Analysis of each individual sample is available in Tables S2-S4 in Multimedia Appendix 1; no significant differences were observed for any nutrient among all samples. Correlations across nutrients within the Irish sample were high, with moderate to high correlations observed for nutrient intake with the Brazilian sample. Poor correlations (r<0.30) were reported for vitamin B12, iodine, vitamin A, and carotene intakes within the Polish sample (Table S3 in Multimedia Appendix 1). Good agreement across individual intake was found across all individual samples for protein, fat, and carbohydrate (as percentage energy); sugar; vitamin D; and calcium (Figures S1-S3 in Multimedia Appendix 1).

Tables 3 and 4 provide an overview of mean population and individual food group intakes across the 2 recall methods. In total, intake of 6 food groups differed significantly across the interviewer-led and self-administered recalls, once adjusted. Of these 6 groups, potatoes and potato dishes and nuts, herbs, and seeds were poorly correlated with high percentage differences observed across recall methods for the nuts, herbs, and seeds group (r=0.56, P<.001, percentage difference=12.18 and r=0.47, P<.001, percentage difference=32.07, respectively; Table 3). Similar trends were identified within individual sample groups, with Brazilian, Irish, and Polish participants overestimating intake of the nuts, herbs, and seeds group by up to 40% and low correlations between methods were observed across each sample for this food group (r=0.39-49; Tables S9-S11 in Multimedia Appendix 1).

Determining cross-classification of food group intakes, all food groups had >80% of respondents in the same or adjacent quartiles, with the *breakfast cereals*; *fish and fish products*; *fruit and fruit juices*; and *sugars, confectionary, and preserves* groups having >95% of records in the same or adjacent quartiles (Table 4). κ coefficients indicate moderate agreement for the *beverages*, *fish and fish products*, and *potatoes and potato dishes* groups; good agreement for the *fruit and fruit juices* and *egg and egg dishes* groups; and excellent agreement for the *breakfast cereals* group where all participants were within the same or adjacent quartile. Less than 15% of participant recalls were ≥2 quartiles apart for all food groups except for the *milk and yoghurts* group (Table 4).

The number of matches, omissions, and intrusions of food items are outlined by sample in Table 5. All food and drink items except water were included in this analysis. Rates of matches, omissions, and intrusions were similar across all 3 samples, with Irish participants least likely to omit food items compared to Brazilian and Polish groups (12% vs 23% and 18%, respectively). Food items most frequently omitted by participants across all groups were sauces and condiments from the *soups, sauces, and miscellaneous* group. Intrusions were most likely to be food items from the *sugars, confectionary, and preserves* group such as biscuits and chocolate.

Figure 3. Difference in protein intake (percentage total energy) reported in the self-administered and interviewer-led recall.



XSL•FO
RenderX

Figure 4. Difference in carbohydrate intake (percentage total energy) reported in the self-administered and interviewer-led recall.

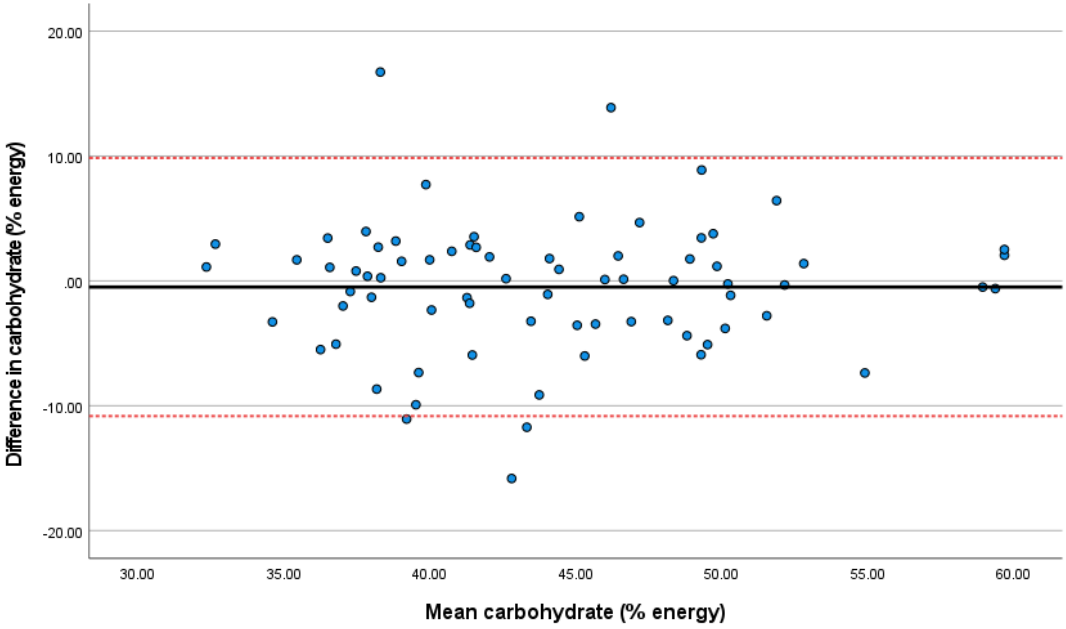
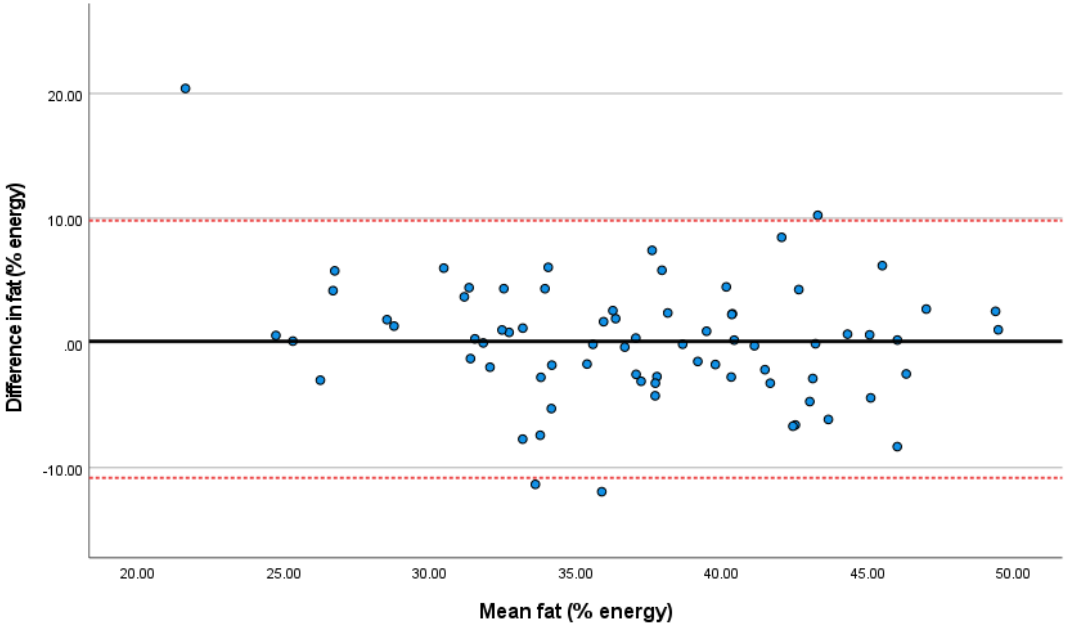


Figure 5. Difference in fat intake (percentage total energy) reported in the self-administered and interviewer-led recall.



XSL•FO
RenderX

Figure 6. Difference in sugar intake (g/day) reported in the self-administered and interviewer-led recall.

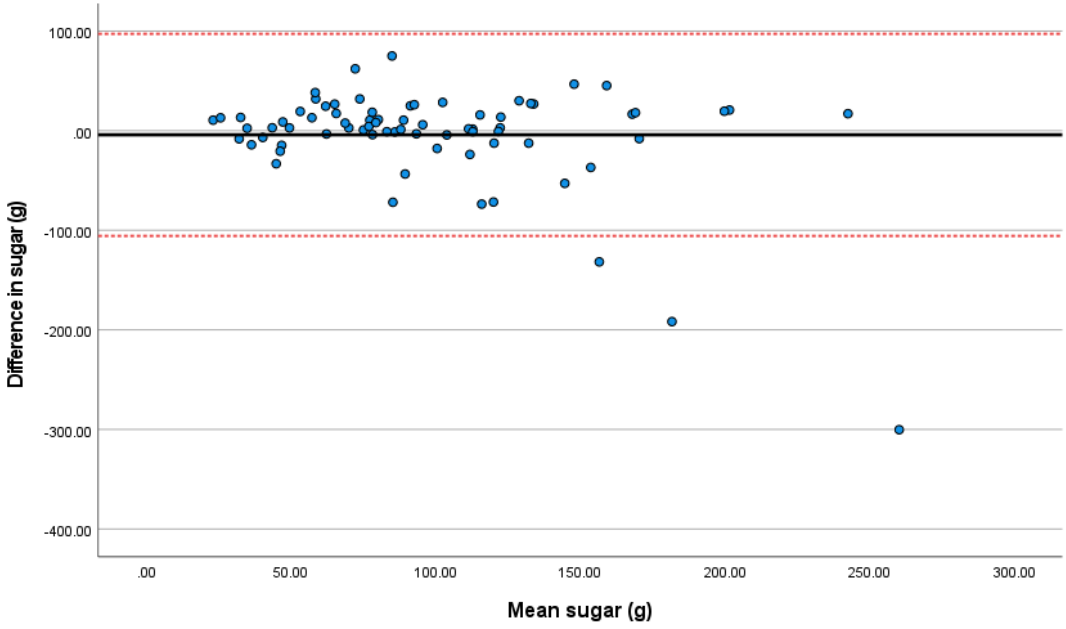
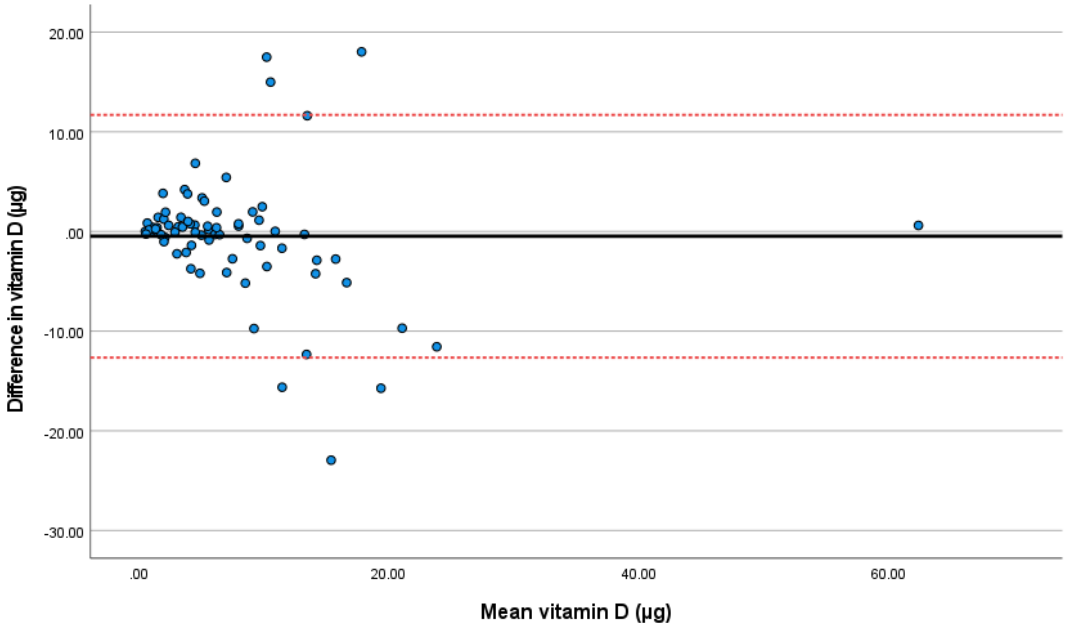
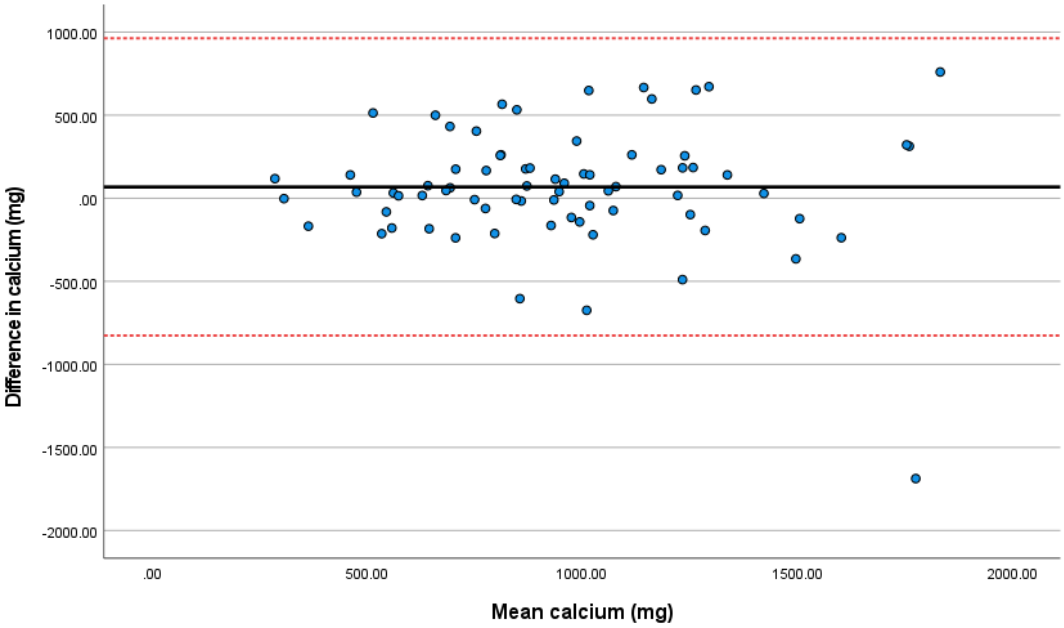


Figure 7. Difference in vitamin D intake (µg/day) reported in the self-administered and interviewer-led recall.



XSL•FO
RenderX

Figure 8. Difference in calcium intake (mg/day) reported in the self-administered and interviewer-led recall.



Discussion

This study has shown that Foodbook24 is a useful method for capturing food group and nutrient intakes in a multinational population living in Ireland. Intakes recorded using Foodbook24 are comparable to those recorded through interviewer-led 24-hour dietary recalls.

Principal Findings

Findings from the qualitative aspects of the Foodbook24 acceptability study demonstrated that most foods consumed by Irish, Brazilian, and Polish adults living in Ireland are available in the Foodbook24 tool and highlighted additional food items that are often consumed by target groups in Ireland. Although this study offered unique insights into the habitual food intake of diverse population subgroups in Ireland, a small sample size of Brazilian and Polish adults was achieved, meaning that the updated Foodbook24 food list may not be wholly representative of the diets of Brazilian and Polish populations in Ireland.

Results from the comparison study indicated similarity in reported food group and nutrient intakes between Foodbook24 and interviewer-led recalls. Comparison studies of similar tools have reported some differences in energy and key micronutrients, including in the original comparison of Foodbook24; however, these were considered small and often insignificant, much like the results presented by Timon et al [20], Foster et al [39], Albar et al [40], and Labonté et al [41]. In line with this study, previous comparison studies of the Foodbook24, Automated Self Administered 24-hour Dietary Assessment Tool (ASA24), and INTAKE24 tools also examined match, omission, and intrusion rates of foods recorded across 2 methods and, similar to the findings reported here, found comparably high match rates (ie, approximately 80%) [20,39,42]. Although statistically significant differences were reported across mean food group intakes among the total sample, the actual differences were minimal and correlations across methods were relatively high within the context of each food group. An exception to this was the *nuts, herbs, and seeds* food group where population and individual intakes indicated low to moderate agreement; overall low consumption of this food group may have exaggerated small differences in intake across methods. Although previous work reported minimal differences in intakes recorded via Foodbook24 and interviewer-led methods for the *nuts, seeds, herbs, and spices* group, this food group has been identified as difficult to quantify using other 24-hour dietary assessment tools, such as INTAKE24 [9,20]. As this food group is typically consumed in small amounts (serving size=30 g), estimating its portion size can be difficult, and the percentage difference may be exaggerated (eg, a 5 g difference between two methods). Altering how the portion size of this food group is recorded, such as recording the number of nuts or teaspoons of seeds, herbs, and spices consumed rather than using weighed portion size food images, may help improve accuracy of reporting for this food group.

Findings in mean food group and nutrient intakes within the Irish, Brazilian, and Polish sample groups are similar to those reported for the total sample. Irish-specific findings align with those previously identified during the initial Foodbook24 comparison study, where high correlations across self-administered and interviewer-led methods were observed [20]. Despite slightly lower correlations among Polish and Brazilian groups, no significant differences were found across intakes and similar intakes across methods indicate suitability of Foodbook24 for assessing dietary intake of Polish and Brazilian adults in Ireland.

Impact of Work

As populations in the developed world continue to diversify, it is essential that tools are developed and validated to be flexible for use by key minority groups of the population. Accurate dietary assessment of diverse population subgroups is lacking both in Ireland and worldwide, and in consequence, there is a paucity of data examining food consumption patterns and food choice influence of differing ethnicities [43]. This in turn

XSL•FO
RenderX

impacts the inclusion of such groups in nutritional research outputs and subsequent public health policy, which may lead to adverse health outcomes among minority groups [44]. Poor consideration of population subgroups when developing dietary assessment tools that are intended to be nationally representative means that many surveys conducted using such tools may not accurately depict intake of different nationalities or ethnicities living in one location. Previous work has developed assessment tools for use across different countries through food list expansion and tool translation, allowing for cross-country comparisons of dietary data [45-47]. However, these developments often cater primarily to the general populations within each country, with minor nationalities being overlooked during the tool development process. Tools that cater for minority nationalities and ethnicities are also important as data collected through national food consumption surveys do not reflect the dietary trends of countries' nationals who have immigrated from elsewhere. When emigrating to a new country, acculturation to the host country's dietary norms and limitations on food availability mean that the food consumption habits of immigrants inevitably change from their traditional diets [43]. However, despite the adoption of some of the host country's dietary norms by immigrants, those from different nationalities often retain certain traditional food items in their new diet, which is particularly common among older generations [48]. Therefore, for dietary assessment to measure food intake of all subgroups of the population, including those of different ethnicities or nationalities, the usability and food lists of web-based dietary assessment tools need to be validated among population subgroups. Furthermore, subsequent diversification of existing food lists is likely to be required. Thompson et al [49] has highlighted the need to consider language, literacy, and content of food lists within dietary assessment tools that are intended to be used among diverse population groups. If used in populations outside those intended (eg, Irish adults in the case of Foodbook24), web-based assessment tools should incorporate different languages, population-specific names for food and drink items, appropriate nutrient composition source, and portion size estimates of these items [49]. In addition, future tool developments should also consider how eating habits (eg, sharing of large plates or traditional dishes) differ across cultures and allow flexibility in how portion size is assessed [7]. By translating the tool into Polish and Portuguese and expanding the food list to include a wide range of ethnic foods, Foodbook24 has been improved for accurate use among diverse populations in Ireland. Advantages of Foodbook24 cited by participants included its ease of use, convenience, and portion size quantification. Although Foodbook24 was cited as being easy to use, some population subgroups (eg, older adults) may benefit from technical support while completing web-based dietary recalls [50]. Web-based platforms such as Zoom, as used in this study, would facilitate this. Although possibly more time consuming for researchers, providing ample support to participants as needed would help improve both retention of participants and accuracy of data collected.

Future Work

While actual measurement of intakes in diverse groups is important, the validation of dietary assessment methods for use in these diverse populations is also essential to ensure dietary patterns are accurately captured. Many comparison and validation studies do not consider differing ethnicities or nationalities within the sample population, with few assessments of other web-based dietary assessment tools worldwide obtaining a rate of participants from ethnic minorities of >10% [9,51-54]. This comparison study was not designed to examine diverse population dietary intakes but rather to test the accuracy of food intake as reported through a self-administered web-based 24-hour dietary recalls by diverse adults in Ireland. Further testing and validation of Foodbook24 among a larger pool of participants would be advantageous. The future use of Foodbook24 could help better understand the habitual diets of different nationalities living in Ireland and inform public health policies aimed at supporting specific population subgroups in meeting healthy eating guidelines and nutrient requirements. Similar tool developments across other countries are essential to understand the food intakes of minority groups globally and ensure that diverse nationalities are included in nutrition monitoring worldwide. With such developments in dietary assessment tools, research into how diets of different nationalities compare and how country of residency impacts the dietary intakes of immigrants is possible and would support public health strategies looking to address the needs of minority groups. It is recognized that inclusion of diverse population subgroups in research is challenging, but more needs to be done to support the participation of differing nationalities in research and provide insight into dietary patterns of the total population [55]. Recruitment and participant retention were major challenges faced in this research and is one that has been faced by many researchers in the past [37,56]. A scoping review by Wieland et al [57] noted community engagement to promote trust is a crucial avenue to explore when hoping to effectively recruit minority groups and future efforts should look to understand cultures and research perspectives of target groups [58].

Strengths and Limitations

In terms from strengths of this research, participants of diverse nationalities were involved in the expansion of Foodbook24 by providing insight into their habitual eating patterns, (ie, indicating additional food items required in Foodbook24) and testing the translated versions of the tool. This is the first study to assess the accuracy of a web-based dietary assessment tool among diverse population groups in Ireland in comparison to traditional assessment methods. Self-administered and interviewer-led recalls presented here reflected the same 24-hour period. All researchers conducting interviewer-led 24-hour recalls were trained and followed a predefined protocol. The same portion size estimates and food composition database was used for both interviewer-led and self-administered recalls so that the full effect of recall method on dietary intakes could be examined.

Despite trying to achieve an equal distribution of Brazilian, Irish, and Polish participants to assess Foodbook24, low numbers were recruited across groups. Just <20% (19/104, 18.27%) of the total sample who completed the comparison study were Polish, and only 13% (4/32) of Brazilian participants were male, meaning that comparisons within and across groups

XSL•FO
RenderX

must be interpreted with caution. Associations identified within Brazilian and Polish samples could be exaggerated or significant differences in intake unidentified; a larger sample size of individual groups would reduce random variation and may reduce larger SD of nutrient and food group intakes observed. Brazilian and Polish participants were unable to complete the interviewer-led recalls in their first language and this may have caused differences in reporting across methods. As with all retrospective dietary intake data collection, this study was reliant on memory of participants and is subject to recall bias. Despite every effort to facilitate participant memory (ie, food prompts and forgotten foods checklist), self-reporting remains flawed. This study focused on developing Foodbook24 for use among non-Irish nationals, who are usually younger in age. Future work may look to develop Foodbook24 among children and additional nationalities growing in Ireland. Recruitment for this study was conducted during the COVID-19 pandemic; thus, in-person engagement with community members proved particularly difficult.

Conclusions

The expanded Foodbook24 food list is comprehensive and generally accepted by Irish, Brazilian, and Polish populations living in Ireland. Foodbook24 is now available in multiple languages, which allows users the option to select which language to record their diet in. Results reported here demonstrate that Foodbook24 is an accurate method of measuring food group and nutrient intakes in a nationally diverse sample of adults living in Ireland when compared to interviewer-led 24-hour dietary recalls. Minimal differences were observed in the food group and nutrient intakes recorded through Foodbook24 across Brazilian, Irish, and Polish groups. Validation involving a larger and more diverse sample as well as biomarker analysis could be completed in the future to further assess the tool. Overall, the expanded Foodbook24 tool allows researchers to accurately identify dietary and meal patterns across different groups living in Ireland, taking diverse language and food choice into consideration.

Acknowledgments

This research was supported by the Food Nutrition Security Cloud (FNS-Cloud). FNS-Cloud has received funding from the European Union's Horizon 2020 Research and Innovation programme (H2020-EU.3.2.2.3; a sustainable and competitive agri-food industry) under grant 863059 [59].

Data Availability

The datasets generated and analyzed during this study are available in the Irish Social-Science Data Archive repository [60]. The updated Foodbook24 tool is freely available on the internet (FoodBook24). Any researcher who is interested in using Foodbook24 for their research should contact author ERG.

Authors' Contributions

LAB and ERG designed protocols for acceptability and comparison studies. GB and SY collected data. GB performed statistical analysis and wrote the manuscript. All authors provided interpretation of results and reviewed final draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results from individual cohorts (ie findings from Irish, Brazilian, and Polish groups separately).
[DOCX File , 1665 KB - ojphi_v17i1e52380_app1.docx]

References

1. European Food Safety Authority. General principles for the collection of national food consumption data in the view of a pan-European dietary survey. EFSA J 2009 Dec 18;7(12):1435. [doi: 10.2903/j.efsa.2009.1435]
2. Amoutzopoulos B, Steer T, Roberts C, Cade JE, Boushey CJ, Collins CE, et al. Traditional methods new technologies - dilemmas for dietary assessment in large-scale nutrition surveys and studies: a report following an international panel discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisbane, 3 September 2015. J Nutr Sci 2018 Apr 02;7:e11 [FREE Full text] [doi: 10.1017/jns.2018.4] [Medline: 29686860]
3. Thompson FE, Subar AF. Dietary assessment methodology. In: Coulston AM, Boushey CJ, Ferruzzi M, editors. Nutrition in the Prevention and Treatment of Disease. 4th Edition. New York, NY: Academic Press; 2022:5-48.
4. Baghlaf K. Traditional dietary assessment tools verses technology-based methods in children and adolescence. An updated review. Prog Nutr 2022;24(1):22 [FREE Full text]
5. Gianfrancesco C, Darwin Z, McGowan L, Smith DM, Haddrill R, Carter M, et al. Exploring the feasibility of use of an online dietary assessment tool (myfood24) in women with gestational diabetes. Nutrients 2018 Aug 23;10(9):1147 [FREE Full text] [doi: 10.3390/nu10091147] [Medline: 30142898]

XSL•FO
RenderX

6. Bonilla C, Brauer P, Royall D, Keller H, Hanning RM, DiCenso A. Use of electronic dietary assessment tools in primary care: an interdisciplinary perspective. BMC Med Inform Decis Mak 2015 Feb 25;15(1):14 [FREE Full text] [doi: 10.1186/s12911-015-0138-6] [Medline: 25886381]
7. Almiron-Roig E, Aitken A, Galloway C, Ellahi B. Dietary assessment in minority ethnic groups: a systematic review of instruments for portion-size estimation in the United Kingdom. Nutr Rev 2017 Mar 01;75(3):188-213 [FREE Full text] [doi: 10.1093/nutrit/nuw058] [Medline: 28340101]
8. Shim J, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014 Jul 22;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
9. Bradley J, Simpson E, Poliakov I, Matthews JN, Olivier P, Adamson AJ, et al. Comparison of INTAKE24 (an online 24-h dietary recall tool) with interviewer-led 24-h recall in 11-24 year-old. Nutrients 2016 Jun 09;8(6):358 [FREE Full text] [doi: 10.3390/nu8060358] [Medline: 27294952]
10. Margaras V. Demographic trends in EU regions. European Parliamentary Research Service. 2019. URL: https://ec.europa.eu/futurium/en/system/files/ged/eprs-briefing-633160-demographic-trends-eu-regions-final.pdf [accessed 2024-04-29]
11. Linz K, Stula S. Demographic change in Europe-an overview. Observatory for Sociopolitical Developements in Europe. 2010. URL: https://tinyurl.com/4k38x6ut [accessed 2024-04-29]
12. Census of population 2016 – profile 8 Irish travellers, ethnicity and religion. Central Statistics Office. URL: https://www.cso.ie/en/releasesandpublications/ep/p-cp8iter/p8iter/p8e/ [accessed 2024-04-29]
13. Redwood S, Gill PS. Under-representation of minority ethnic groups in research — call for action. Br J Gen Pract 2013 Jul 01;63(612):342-343. [doi: 10.3399/bjgp13x668456]
14. Etti M, Fofie H, Razai M, Crawshaw AF, Hargreaves S, Goldsmith LP. Ethnic minority and migrant underrepresentation in COVID-19 research: causes and solutions. EClinicalMedicine 2021 Jun;36:100903 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100903] [Medline: 34095791]
15. Leung G, Stanner S. Diets of minority ethnic groups in the UK: influence on chronic disease risk and implications for prevention. Nutr Bull 2011;36(2):161-198. [doi: 10.1111/j.1467-3010.2011.01889.x]
16. Zoellner J, Anderson J, Gould SM. Development and formative evaluation of a bilingual interactive multimedia dietary assessment tool. J Ext 2006;44:8 [FREE Full text]
17. National Nutrition Survey menuCH 2014-2015. Center for Primary Care and Public Health (Unisanté) UoL, Switzerland (Unisanté), Swiss Federal Food Safety and Veterinary Office (FSVO). URL: https://tinyurl.com/3dcrhzjz [accessed 2024-04-29]
18. Evans K, Hennessy Á, Walton J, Timon C, Gibney E, Flynn A. Development and evaluation of a concise food list for use in a web-based 24-h dietary recall tool. J Nutr Sci 2017 Aug 29;6:e46 [FREE Full text] [doi: 10.1017/jns.2017.49] [Medline: 29152250]
19. Timon CM, Blain RJ, McNulty B, Kehoe L, Evans K, Walton J, et al. The development, validation, and user evaluation of Foodbook24: a web-based dietary assessment tool developed for the Irish adult population. J Med Internet Res 2017 May 11;19(5):e158 [FREE Full text] [doi: 10.2196/jmir.6407] [Medline: 28495662]
20. Timon CM, Evans K, Kehoe L, Blain RJ, Flynn A, Gibney ER, et al. Comparison of a web-based 24-h dietary recall tool (Foodbook24) to an interviewer-led 24-h dietary recall. Nutrients 2017 Apr 25;9(5):425 [FREE Full text] [doi: 10.3390/nu9050425] [Medline: 28441358]
21. Rodrigues PR, Monteiro LS, Cunha DB, Sichieri R, Pereira RA. Adult food consumption by household composition: an analysis of the first National Dietary Survey, Brazil, 2008-2009. Public Health Nutr 2020 Feb 02;23(2):193-201 [FREE Full text] [doi: 10.1017/S1368980019002374] [Medline: 31475661]
22. Pereira RA, Souza AM, Duffey KJ, Sichieri R, Popkin BM. Beverage consumption in Brazil: results from the first National Dietary Survey. Public Health Nutr 2015 May 27;18(7):1164-1172 [FREE Full text] [doi: 10.1017/S1368980014001657] [Medline: 25158687]
23. Levy-Costa RB, Sichieri R, Pontes Ndos S, Monteiro CA. [Household food availability in Brazil: distribution and trends (1974-2003)]. Rev Saude Publica 2005 Aug;39(4):530-540 [FREE Full text] [doi: 10.1590/s0034-89102005000400003] [Medline: 16113900]
24. Bezerra IN, Goldman J, Rhodes DG, Hoy MK, de Moura Souza A, Chester DN, et al. Difference in adult food group intake by sex and age groups comparing Brazil and United States nationwide surveys. Nutr J 2014 Jul 21;13(1):74 [FREE Full text] [doi: 10.1186/1475-2891-13-74] [Medline: 25047421]
25. Monteiro LS, Hassan BK, Estima CC, Souza AM, Verly EJ, Sichieri R, et al. Food consumption according to the days of the week - national food survey, 2008-2009. Rev Saude Publica 2017 Oct 05;51:93 [FREE Full text] [doi: 10.11606/S1518-8787.2017051006053] [Medline: 29020121]
26. Szponar L, Sekula W, Nelson M, Weisell RC. The household food consumption and anthropometric survey in Poland. Public Health Nutr 2001 Oct;4(5B):1183-1186. [Medline: 11924945]
27. Babicz-Zielinska E. Food preferences among the Polish young adults. Food Qual Prefer 1999 Mar;10(2):139-145. [doi: 10.1016/S0950-3293(99)00005-1]

XSL•FO
RenderX

28. Sekula W, Nelson M, Figurska K, Oltarzewski M, Weisell R, Szponar L. Comparison between household budget survey and 24-hour recall data in a nationally representative sample of Polish households. Public Health Nutr 2005 Jun 02;8(4):430-439. [doi: 10.1079/phn2004695] [Medline: 15975190]
29. Elmadfa I. European nutrition and health report 2009: food consumption in adults on the basis of dietary surveys in European countries. Karger Publishers. 2009. URL: https://tinyurl.com/bp58mxf2 [accessed 2024-04-29]
30. McCance RA, Widdowson EM. McCance and Widdowson's Composition of Foods, 7th edition. London, UK: Royal Society of Chemistry; 2014.
31. Dataset records for national food and nutrition institute (Poland). Institute for Health Metrics and Evaluation. URL: https://ghdx.healthdata.org/organizations/national-food-and-nutrition-institute-poland [accessed 2024-04-29]
32. Coelho KS, Bistriche Giuntini EB, Grande F, da Silva Dias J, Purgatto E, de Melo Franco B, et al. 12th IFDC 2017 special issue – Brazilian Food Composition Table (TBCA): development and functionalities of the online version. J Food Compost Anal 2019 Dec;84:103287. [doi: 10.1016/j.jfca.2019.103287]
33. Crawley H. Food Portion Sizes. London, UK: H.M. Stationery Office & Ministry of Agriculture, Fisheries and Food; 1994.
34. Budig K, Diez J, Conde P, Sastre M, Hernán M, Franco M. Photovoice and empowerment: evaluating the transformative potential of a participatory action research project. BMC Public Health 2018 Apr 02;18(1):432 [FREE Full text] [doi: 10.1186/s12889-018-5335-7] [Medline: 29609576]
35. Díez J, Conde P, Sandin M, Urtasun M, López R, Carrero JL, et al. Understanding the local food environment: a participatory photovoice project in a low-income area in Madrid, Spain. Health Place 2017 Jan;43:95-103. [doi: 10.1016/j.healthplace.2016.11.012] [Medline: 27984767]
36. Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, Rumpler WV, et al. The US Department of Agriculture automated multiple-pass method reduces bias in the collection of energy intakes. Am J Clin Nutr 2008 Aug;88(2):324-332 [FREE Full text] [doi: 10.1093/ajcn/88.2.324] [Medline: 18689367]
37. The calculation of Bonferroni-adjusted p-values. IBM Corp. URL: https://tinyurl.com/35v9zaa8 [accessed 2024-04-29]
38. Henry CJ. Basal metabolic rate studies in humans: measurement and development of new equations. Public Health Nutr 2005 Oct 02;8(7A):1133-1152. [doi: 10.1079/phn2005801] [Medline: 16277825]
39. Foster E, Delve J, Simpson E, Breininger SP. Comparison study: INTAKE24 vs interviewer led recall final report. Food Standards Agency. 2014. URL: https://ndns.intake24.org/assets/papers/Intake24-Comparison-report.pdf [accessed 2024-04-29]
40. Albar SA, Alwan NA, Evans CEL, Greenwood DC, Cade JE. Agreement between an online dietary assessment tool (myfood24) and an interviewer-administered 24-h dietary recall in British adolescents aged 11-18 years. Br J Nutr 2016 May 15;115(9):1678-1686 [FREE Full text] [doi: 10.1017/S0007114516000593] [Medline: 26975650]
41. Labonté M, Cyr A, Baril-Gravel L, Royer MM, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. Eur J Clin Nutr 2012 Feb 21;66(2):166-173. [doi: 10.1038/ejcn.2011.163] [Medline: 21934698]
42. Kirkpatrick SI, Subar AF, Douglass D, Zimmerman TP, Thompson FE, Kahle LL, et al. Performance of the automated self-administered 24-hour recall relative to a measure of true intakes and to an interviewer-administered 24-h recall. Am J Clin Nutr 2014 Jul;100(1):233-240 [FREE Full text] [doi: 10.3945/ajcn.114.083238] [Medline: 24787491]
43. Bennett G, Bardon LA, Gibney ER. A comparison of dietary patterns and factors influencing food choice among ethnic groups living in one locality: a systematic review. Nutrients 2022 Feb 23;14(5):941 [FREE Full text] [doi: 10.3390/nu14050941] [Medline: 35267916]
44. Jackson CS, Gracia JN. Addressing health and health-care disparities: the role of a diverse workforce and the social determinants of health. Public Health Rep 2014 Jan 01;129(1_suppl2):57-61. [doi: 10.1177/00333549141291s211]
45. Comparison of ASA24® versions. National Cancer Institute: Division of Cancer Control & Population Science. URL: https://epi.grants.cancer.gov/asa24/comparison.html [accessed 2024-04-29]
46. FAQs: what languages is myfood24 available in? myfood24. URL: https://www.myfood24.org/about-myfood24 [accessed 2024-04-29]
47. Verger EO, Armstrong P, Nielsen T, Chakaroun R, Aron-Wisnewsky J, Gøbel RJ, MetaCardis Consortium. Dietary assessment in the MetaCardis study: development and relative validity of an online food frequency questionnaire. J Acad Nutr Diet 2017 Jun;117(6):878-888. [doi: 10.1016/j.jand.2016.10.030] [Medline: 28024800]
48. Satia-Abouta J, Patterson RE, Neuhouser ML, Elder J. Dietary acculturation: applications to nutrition research and dietetics. J Am Diet Assoc 2002 Aug;102(8):1105-1118. [doi: 10.1016/s0002-8223(02)90247-6] [Medline: 12171455]
49. Thompson FE, Dixit-Joshi S, Potischman N, Dodd KW, Kirkpatrick S, Kushi L, et al. Comparison of interviewer-administered and automated self-administered 24-hour dietary recalls in 3 diverse integrated health systems. Am J Epidemiol 2015 Jun 15;181(12):970-978 [FREE Full text] [doi: 10.1093/aje/kwu467] [Medline: 25964261]
50. Timon CM, Astell AJ, Hwang F, Adlam TD, Smith T, Maclean L, et al. The validation of a computer-based food record for older adults: the Novel Assessment of Nutrition and Ageing (NANA) method. Br J Nutr 2015 Feb 28;113(4):654-664 [FREE Full text] [doi: 10.1017/S0007114514003808] [Medline: 25630436]
51. Wark PA, Hardie LJ, Frost GS, Alwan NA, Carter M, Elliott P, et al. Validity of an online 24-h recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews. BMC Med 2018 Aug 09;16(1):136 [FREE Full text] [doi: 10.1186/s12916-018-1113-8] [Medline: 30089491]

XSL•FO
RenderX

52. Haftenberger M, Heuer T, Heidemann C, Kube F, Krems C, Mensink GB. Relative validation of a food frequency questionnaire for national health and nutrition monitoring. Nutr J 2010 Sep 14;9(1):36 [FREE Full text] [doi: 10.1186/1475-2891-9-36] [Medline: 20840739]
53. Schröder H, Covas MI, Marrugat J, Vila J, Pena A, Alcántara M, et al. Use of a three-day estimated food record, a 72-hour recall and a food-frequency questionnaire for dietary assessment in a Mediterranean Spanish population. Clin Nutr 2001 Oct;20(5):429-437. [doi: 10.1054/clnu.2001.0460] [Medline: 11534938]
54. Buscemi S, Rosafio G, Vasto S, Massenti FM, Grosso G, Galvano F, et al. Validation of a food frequency questionnaire for use in Italian adults living in Sicily. Int J Food Sci Nutr 2015 Apr 22;66(4):426-438. [doi: 10.3109/09637486.2015.1025718] [Medline: 25830946]
55. Willis A, Isaacs T, Khunti K. Improving diversity in research and trial participation: the challenges of language. Lancet Public Health 2021 Jul;6(7):e445-e446 [FREE Full text] [doi: 10.1016/S2468-2667(21)00100-6] [Medline: 34174997]
56. Rooney LK, Bhopal R, Halani L, Levy ML, Partridge MR, Netuveli G, et al. Promoting recruitment of minority ethnic groups into research: qualitative study exploring the views of South Asian people with asthma. J Public Health (Oxf) 2011 Dec;33(4):604-615. [doi: 10.1093/pubmed/fdq100] [Medline: 21228023]
57. Wieland ML, Njeru JW, Alahdab F, Doubeni CA, Sia IG. Community-engaged approaches for minority recruitment into clinical research: a scoping review of the literature. Mayo Clin Proc 2021 Mar;96(3):733-743. [doi: 10.1016/j.mayocp.2020.03.028] [Medline: 33004216]
58. Levkoff S, Sanchez H. Lessons learned about minority recruitment and retention from the Centers on Minority Aging and Health Promotion. Gerontologist 2003 Feb 01;43(1):18-26. [doi: 10.1093/geront/43.1.18] [Medline: 12604742]
59. Welcome to FNS-cloud. Food Nutrition Security Cloud (FNS-Cloud). URL: https://www.fns-cloud.eu/ [accessed 2024-04-29]
60. Irish social science data archive. University College Dublin. URL: https://www.ucd.ie/issda/ [accessed 2024-04-29]

Abbreviations

CoFID: Composition of Food Integrated Database
UCD: University College Dublin

Edited by E Mensah; submitted 01.09.23; peer-reviewed by T Baranowski, C Collins; comments to author 12.04.24; revised version received 27.05.24; accepted 09.10.24; published 07.02.25.

Please cite as:
Bennett G, Yang S, Bardon LA, Timon CM, Gibney ER
Expansion and Assessment of a Web-Based 24-Hour Dietary Recall Tool, Foodbook24, for Use Among Diverse Populations Living in Ireland: Comparative Analysis
Online J Public Health Inform 2025;17:e52380
URL: https://ojphi.jmir.org/2025/1/e52380
doi:10.2196/52380
PMID:39919284

©Grace Bennett, Shuhua Yang, Laura A Bardon, Claire M Timon, Eileen R Gibney. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Disparities in Tobacco Use, Binge Drinking, and Injection Drug Use Among Sexual and Gender Minoritized People Recruited Online in San Francisco, California: Survey Results

Sean Arayasirikul[1], PhD; Jarett Maycott[2], MPH; Arianna Contestable[1], BA

[1]The Legacy Center, Joe C Wen School of Population & Public Health, University of California, Irvine, Irvine, CA, United States

[2]San Francisco Department of Public Health, San Francisco, CA, United States

Corresponding Author:
Sean Arayasirikul, PhD
The Legacy Center
Joe C Wen School of Population & Public Health
University of California, Irvine
856 Health Sciences Quad
Irvine, CA, 92697-3957
United States
Phone: 1 949 824 2680
Email: sean.arayasirikul@uci.edu

Abstract

This research letter describes substance use disparities among online, help-seeking, sexual and gender minoritized people in San Francisco. Our findings emphasize the importance of strengthening public health practice by leveraging digital and online methods to reach communities.

(Online J Public Health Inform 2025;17:e81982) doi:10.2196/81982

KEYWORDS

substance use; health disparities; sexual and gender minorities; community health; social media

Introduction

The complexity of substance use as a public health issue is multifaceted. Compared to the general population, people who use substances are more likely to have hospital and emergency department visits, often presenting with an immediate need for treatment (eg, overdose) or being a source of harm to others (eg, driving under the influence) [1]. Research has identified people who use substances as a population that utilizes greater health care services at higher costs [2]. In response to this strain on the public health care system, specialized services such as sobering centers and supervised injection facilities have become critical harm reduction assets [3,4]. While the public health infrastructure and tools to prevent substance use disorder has improved to address the opioid epidemic, people who use substances experience substance use stigma as a barrier to accessing substance use services [5]. Substance use impacts diverse marginalized communities, especially sexual and gender minoritized (SGM) communities. SGM individuals are more likely to use substances compared to cisgender heterosexual individuals [6,7].

Methods

Study Design

We conducted a cross-sectional analysis of 409 adult men who have sex with men and transgender women and gender expansive (TGE) in San Francisco. Participants interested in receiving digital navigation services for substance use and related health topics (eg, mental health, HIV) were recruited using advertisements on Facebook, Instagram, and Grindr in 2022-2024. We analyzed the sociodemographic characteristics. Race/ethnicity was recoded to a dichotomous variable with Black, indigenous, and people of color (BIPOC) people=1 and White=0. We assessed participants' recent substance use behaviors by asking how many days in the last 30 days, if any, did they use the following substances [8]: tobacco, vaping products, binge drinking, marijuana, prescription opioids not prescribed to you (eg, pain relievers), nonprescription opioids (eg, heroin, fentanyl), other prescription drugs (for nonmedical purposes), other illegal substances, and injected any drugs. Responses for any recent use were recoded and dichotomized for each substance (Yes/No). We describe the sample and

XSL•FO
RenderX

characterize recent substance use by using univariate statistics. Logistic regression was used to identify associations between BIPOC status, TGE status, and HIV status and each substance use behavior. We built logistic regression models for each substance use outcome, adjusted for age, sexual orientation, and housing status. Analyses were performed using Stata (version 17; StataCorp LLC).

Ethical Considerations

The study protocol was approved by the University of California, San Francisco institutional review board (20-33169). Participants provided signed informed consent and were remunerated US $30 (gift card) for completing the assessment.

Results

The sample was racially/ethnically diverse (Table 1), with 58.68% (240/409) of the sample comprised of BIPOC participants. One in 5 participants (82/409, 20.05%) were TGE, and nearly a third (131/409, 32.03%) were people living with HIV. Marijuana (220/409, 53.79%) and other illegal substance use (227/409, 55.50%) were the most reported, followed by binge drinking (160/409, 39.12%), tobacco use (155/409, 37.90%), and vaping (121/409, 29.58%). Prescription and nonprescription opioid use (such as heroin or fentanyl) was reported by a small proportion of the sample (6/409, 1.47% and 11/409, 2.69%, respectively), while other prescription drug use was 11.74% (48/409), and 10.51% (43/409) reported injection drug use. BIPOC participants had higher odds of tobacco use compared to White participants (adjusted odds ratio [aOR]=1.77, 95% CI 1.13-2.80). TGE participants also had significantly greater odds of tobacco use compared to cisgender men (aOR=3.18, 95% CI 1.70-6.12). People living with HIV had significantly higher odds of both tobacco use (aOR=1.78, 95% CI 1.10-2.89) and injection drug use (aOR=3.14, 95% CI 1.53-6.69) compared to counterparts not living with HIV (Table 2).

XSL•FO
RenderX

Table 1. Demographic characteristics and substance use behaviors among online help-seeking sexual and gender minoritized people in San Francisco, California, in 2021-2024 (N=409).

	Values, n (%)
Demographics	
Age (y)	
18-29	103 (25.18)
30-39	111 (27.14)
40-49	77 (18.83)
50+	118 (28.85)
Race/ethnicity	
White	169 (41.32)
Latino/a/x/e	113 (27.63)
Asian, Pacific Islander, and Native Hawaiian	50 (12.22)
Black	38 (9.29)
Multiracial/other	39 (9.54)
Gender	
Cisgender man	327 (79.95)
Transgender women or gender expansive	82 (20.05)
Sexual orientation	
Bisexual	47 (11.49)
Gay/lesbian	282 (68.95)
Other	48 (11.74)
Straight/heterosexual	32 (7.82)
HIV status	
Person not living with HIV	278 (67.97)
Person living with HIV	131 (32.03)
Housing stability	
Stable	337 (82.40)
Unstable	72 (17.60)
Socioeconomic status	
Above Federal Poverty Line	337 (82.40)
Below Federal Poverty Line	72 (17.60)
Substance use behaviors (past 30 days)	
Tobacco	
No	254 (62.10)
Yes	155 (39.90)
Vaping	
No	288 (70.42)
Yes	121 (29.58)
Binge drinking	
No	249 (60.88)
Yes	160 (39.12)
Marijuana	
No	189 (46.21)

XSL•FO
RenderX

	Values, n (%)
Yes	220 (53.79)
Prescription opioids	
No	403 (98.53)
Yes	6 (1.47)
Nonprescription opioids	
No	398 (97.31)
Yes	11 (2.69)
Other prescription drugs	
No	361 (88.26)
Yes	48 (11.74)
Other illegal substances	
No	182 (44.50)
Yes	227 (55.50)
Injection drug use	
No	366 (89.49)
Yes	43 (10.51)

Table 2. Associations between substance use and social inequity among online help-seeking sexual and gender minoritized people in San Francisco, California, 2021-2024 (N=409).

	BIPOC[a], aOR[b] (95% CI)	TGE[c], aOR (95% CI)	PLWH[d], aOR (95% CI)
Tobacco	1.77 (1.13-2.80)	3.18 (1.70-6.12)	1.78 (1.10-2.89)
Vaping	0.98 (0.61-1.58)	1.05 (0.54-2.00)	1.16 (0.68-1.98)
Binge drinking	0.74 (0.48-1.15)	0.67 (0.34-1.30)	0.57 (0.35-0.93)
Marijuana	1.10 (0.72-1.67)	1.34 (0.73-2.52)	0.92 (0.59-1.46)
Illicit opioids	2.82 (0.68-19.20)	4.14 (0.94-17.30)	0.52 (0.07-2.32)
Prescription opioids	0.68 (0.11-5.60)	0.67 (0.06-5.53)	0.47 (0.05-3.21)
Other prescription drugs	1.03 (0.55-1.98)	0.85 (0.31-2.11)	0.95 (0.46-1.89)
Other illegal drugs	1.10 (0.72-1.69)	0.59 (0.31-1.11)	1.43 (0.89-2.30)
Injection drug use	1.03 (0.51-2.11)	1.15 (0.39-3.17)	3.14 (1.53-6.69)

[a]BIPOC: Black, indigenous, and people of color.

[b]aOR: adjusted odds ratio.

[c]TGE: transgender women and gender expansive.

[d]PLWH: people living with HIV.

Discussion

Our findings characterize the diverse substance use behavior profiles of SGM people seeking help online in San Francisco. Tobacco use showed a disparity across all groups, likely due to exposure to social stress and discrimination [9]. We also found that people living with HIV had significantly greater odds of injection drug use—a finding that may reflect the syndemic relationship between substance use and HIV as well as persistent structural vulnerabilities. The generalizability of this study is limited by its use of online recruitment and focus on San Francisco, a location that has successfully adopted harm reduction policies [10]. Despite this, community-based studies can supplement nationally representative, publicly available datasets that may undersample SGM and BIPOC communities. While much of substance use research has examined illicit substance use among SGM communities, our findings are a sober reminder that the need for tobacco cessation is also great. The implications of these findings may include bolstering local public health efforts to utilize digital methods to conduct outreach and deliver brief behavioral interventions directly to communities actively seeking help.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank all participants in the study. This work was funded by the Substance Abuse and Mental Health Service Administration (award H79SP082077).

Data Availability

The datasets generated or analyzed during this study are not publicly available as they contain details that could be used to identify participants but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Gardner RL, Baier RR, Cooper EL, Clements EE, Belanger E. Interventions to reduce hospital and emergency department utilization among people with alcohol and substance use disorders: a scoping review. Med Care 2022 Feb 01;60(2):164-177. [doi: 10.1097/MLR.0000000000001676] [Medline: 34908009]
2. Berkman ND, Chang E, Seibert J, et al. Management of high-need, high-cost patients: a best fit framework synthesis, realist review, and systematic review. AHRQ Comparative Effectiveness Reviews 2021 Oct:1. [Medline: 34780127]
3. Marshall B, McGlynn E, King A. Sobering centers, emergency medical services, and emergency departments: a review of the literature. Am J Emerg Med 2021 Feb;40:37-40. [doi: 10.1016/j.ajem.2020.11.031] [Medline: 33340876]
4. Levengood TW, Yoon GH, Davoust MJ, Ogden SN, Marshall BDL, Cahill SR, et al. Supervised injection facilities as harm reduction: a systematic review. Am J Prev Med 2021 Nov;61(5):738-749 [FREE Full text] [doi: 10.1016/j.amepre.2021.04.017] [Medline: 34218964]
5. Cernasev A, Hohmeier KC, Frederick K, Jasmin H, Gatwood J. A systematic literature review of patient perspectives of barriers and facilitators to access, adherence, stigma, and persistence to treatment for substance use disorder. Explor Res Clin Soc Pharm 2021 Jun;2:100029 [FREE Full text] [doi: 10.1016/j.rcsop.2021.100029] [Medline: 35481114]
6. Freitag TM, Chen-Sankey JC, Duarte DA, Ramsey MW, Choi K. Variations in substance use and disorders among sexual minorities by race/ethnicity. Subst Use Misuse 2021;56(7):921-928 [FREE Full text] [doi: 10.1080/10826084.2021.1899225] [Medline: 33821743]
7. Hughto JMW, Quinn EK, Dunbar MS, Rose AJ, Shireman TI, Jasuja GK. Prevalence and co-occurrence of alcohol, nicotine, and other substance use disorder diagnoses among US transgender and cisgender adults. JAMA Netw Open 2021 Feb 01;4(2):e2036512 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.36512] [Medline: 33538824]
8. Substance Abuse and Mental Health Services Administration. National survey on drug use and health: summary of methodological studies, 1971-2014. Substance Abuse and Mental Health Services Administration 2014:1. [Medline: 30199189]
9. Rosenthal SR, Gately KA, Archibald N, Baker AB, Dawes MP, Swanberg JE. Substance misuse among sexual and gender minorities: the role of everyday discrimination and identity. Subst Use Misuse 2023;58(14):1874-1882. [doi: 10.1080/10826084.2023.2257309] [Medline: 37707795]
10. Watters JK, Estilo MJ, Clark GL, Lorvick J. Syringe and needle exchange as HIV/AIDS prevention for injection drug users. JAMA 1994 Jan 12;271(2):115-120. [Medline: 8264065]

Abbreviations

aOR: adjusted odds ratio
BIPOC: Black, indigenous, and people of color
SGM: sexual and gender minoritized
TGE: transgender women and gender expansive

XSL•FO
RenderX

Edited by E Mensah; submitted 06.08.25; peer-reviewed by Y Tran, S Sahai; comments to author 26.09.25; revised version received 27.09.25; accepted 04.11.25; published 17.11.25.

Please cite as:
Arayasirikul S, Maycott J, Contestable A
Disparities in Tobacco Use, Binge Drinking, and Injection Drug Use Among Sexual and Gender Minoritized People Recruited Online in San Francisco, California: Survey Results
Online J Public Health Inform 2025;17:e81982
URL: https://ojphi.jmir.org/2025/1/e81982
doi:10.2196/81982
PMID:41248490

©Sean Arayasirikul, Jarett Maycott, Arianna Contestable. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 17.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

One Health Index Calculator for India Using Empirical Methods for Policy Stewardship: Development and Usability Study

Saveetha Meganathan*, PhD; Arpit Katiyar*, MSc; Esha Srivastava, MPH; Rakesh Kumar Mishra, PhD

Tata Institute for Genetics and Society, Bengaluru, India

*these authors contributed equally

Corresponding Author:
Rakesh Kumar Mishra, PhD
Tata Institute for Genetics and Society, Bengaluru, India

Abstract

Background: *One Health* is a collaborative approach that can be used to evaluate and enhance the fields of human, animal, and environmental health and to emphasize their sectoral interconnectedness. Empirical evaluation of the *One Health* performance of a country in the form of an index, provides direction for actionable interventions such as targeted funding, prioritized resource allocation, rigorous data management, and evidence-based policy decisions. These efforts, along with public engagement and awareness on disease management; environmental degradation, and preparedness toward disease outbreaks, contribute to strengthening global health security. Thus, developing a One Health Index (OHI) calculator for India is a significant step toward evidence-based *One Health* governance in the context of low-and middle-income countries.

Objective: This study aimed to (1) develop a OHI Calculator for India using efficient and user-friendly weighting methods and demonstrate the calculation of the OHI; (2) develop India-specific datasets through secondary data collection from reliable data sources; and (3) determine data gaps for policy stewardship.

Methods: We proposed a OHI calculator to measure the OHI from an Indian context by adopting the Global One Health Index framework that comprises 3 categories: 13 key indicators, 57 indicators, and 216 subindicators. Secondary data collection was conducted to create a dataset for specific to India from reliable sources. For measuring OHI, we demonstrated two mathematical weighting methods: an efficient expert-based rating using fuzzy extent analysis and a modified entropy-based weightage method.

Results: We demonstrate the step-by-step OHI calculation by determining indicator scores using both fuzzy extent analysis and modified entropy-based weightage method. Through secondary data collection an India-specific dataset was created using reliable sources. For the datasets from India, data for 156/216 subindicators were available, while that for the remaining 60 indicators were unavailable. Further, a pilot correlation analysis was performed between 20 indicator scores and relevant budget allocations for the years 2022 - 2023, 2023 - 2024, and 2024 - 2025. It was found that increases in the budget allocation across consecutive years improved indicator scores or better performance and vice versa.

Conclusions: The demonstrated OHI calculator has the potential to serve as a governance tool while promoting data transparency and ethical data management. There is a need for a collaborative data federation approach to resolve data gaps, including incomplete, missing, or unavailable data. Further, the correlation analysis between budgetary allocation and performance of indicators provides empirical evidence for policymakers to improve intersectoral communication, multistakeholder engagement, concerted interventions, and informed policy decisions for resource allocation.

(Online J Public Health Inform 2025;17:e65039) doi:10.2196/65039

KEYWORDS

one health; one health index calculator; fuzzy extent analysis; modified entropy-based weightage method; policy stewardship; India; low and middle income countries; LMIC; Index; indicator; secondary data collection

Introduction

One Health is a participatory, collaborative approach to enhancing the health of people, animals, and ecosystems over time. It recognizes the interdependence of the health of humans, domestic and wild animals, plants, and the larger environment. While health, food sources, water, energy, and the environment are broad topics with sector-specific concerns, cross-sectoral and cross-disciplinary collaboration helps protect human health, address health challenges such as the emergence of infectious diseases and antimicrobial resistance, ensure food safety, and promote the health and integrity of ecosystems. The *One Health* approach has the potential to address the complete spectrum of disease control, from prevention to detection, readiness, response, and management, while also contributing to global health security by considering the multisectoral interconnectedness and their mutual impact [1].

XSL•FO
RenderX

The term 'One Health' evolved during the outbreaks of SARS in 2003 - 2004 and H5N1 influenza (ie, bird flu), highlighting the interconnectedness of human, animal, and environmental health. The 'Manhattan Principles' underscored this link, recognizing the need for collaborative approaches in global disease prevention [2]. These outbreaks emphasized the risks posed by unknown pathogens from wildlife, which led to the development of more effective alert and response systems. Global cooperation involving United Nations, World Health Organization, Food and Agriculture Organization, World Organization for Animal Health, United Nations Children's Fund, and the World Bank addressed the H5N1 outbreak, with the International Ministerial Conference on Avian and Pandemic Influenza playing a key role [3]. The primary drivers of the emergence of novel zoonotic infectious diseases include human activities, changes to ecosystems, land use, agriculture intensification, urbanization, international travel, and trade over the past three decades. These diseases, predominantly originating in wildlife, pose significant public health risks. The *One Health* approach is pivotal for the prevention, monitoring, and surveillance of zoonotic and emerging infectious diseases, sustaining food security, and combating antimicrobial resistance, all of which affect human, animal, and environmental health. Subsequently, collaborative monitoring systems have become essential for effectively managing pandemics and outbreaks, given the multitude of health crises in the last decade [3].

Developing a *One Health* Index (OHI) framework can contribute to the implementation of the *One Health* approach by focusing on intersectoral collaborations and the corresponding datasets. The value of such an index extends beyond compiling data; it has the potential to revolutionize our understanding, management, and response to the complex web of factors that influence global health. A OHI simplifies health data, making it accessible to diverse stakeholders including government functionaries, policymakers, health care providers and the public [4]. It condenses a huge volume of information into a single quantifiable value. Government and health care organizations can use these health indices to make informed policy decisions regarding the allocation of resources and implementation of public health interventions [5]. Further, these indices facilitate the monitoring of health trends in a given region or country, enabling the assessment of the effectiveness of public health campaigns and improvements in health care systems and policy frameworks [6,7].

In this context, the Global One Health Index (GOHI) serves as an empirical tool for the systematic assessment of *One Health* scores for more than 200 countries and territories worldwide [8]. The GOHI framework consists of three major categories—extrinsic driver index, core driver index, and intrinsic driver index—which are further categorized into 13 key-indicators, 57 indicators, and 216 subindicators. This structure makes it possible to quantify the *One Health* Index for a country by mapping the multisectoral variables that contribute to the well-being of humans, animals, and the environment and their mutual interactions [8]. Additionally, the GOHI framework accounts for the recruitment of 29 domain experts to attribute weightage to the indicators based on their sectoral experience, further demonstrating the need for cross-sectoral communication and concerted efforts to achieve better *One Health* outcomes for a country [8].

In this study, we proposed to adapt and contextualize the *One Health* Index calculator for India based on the GOHI framework. We demonstrated the calculation of indicator scores using two mathematical weighting methods: (1) the efficient expert-based fuzzy extent analysis (FEA) method that requires recruitment of domain experts, and (2) the modified entropy-based weightage method (MEWM), which replaces expert-based weightage calculation with mathematical formula-based calculations. A country-specific database for India was developed through secondary data collection. The two weighting methods—FEA and MEWM—were demonstrated using the India-specific dataset to generate indicator scores, subject to data availability. In addition, we correlated sectoral budget allocations over the last two financial years with corresponding sectoral indicator scores. This analysis serves as an actionable pointer for policy makers, federal and state governments to support decision-making toward budgetary allocations.

Methods

The OHI calculator for India is aimed to serve as a public health tool that addresses the need for multisectoral collaboration for data access and understanding the impact of sectoral performance, thereby improving the countrywide *One Health* performance. This calculator consists of the following steps:

Indicator Selection

To demonstrate the calculation of the OHI for India, we adopted three categories: 13 key indicators, 57 indicators, and 216 subindicators from the GOHI [8] (Multimedia Appendix 1).

Database Building

For the demonstration of the OHI calculator for India, national-level data collection was conducted using secondary data sources such as the Press Information Bureau (Government of India), Department of Agriculture and Farmers Welfare (Government of India), World Bank, Food and Agriculture Organization of the United Nations, Yale Environmental Performance Index (Yale University), Our World in Data, The Global Economy, Statista, IndiaStat, Knoema, and other relevant databases (Multimedia Appendix 2). Additionally, to correlate the sectoral budgets over the last two financial years with their respective indicator scores, the budget datasets were obtained from the Indian Union Budget documents for the financial years 2022 - 2023, 2023 - 2024, and 2024 - 2025 (Multimedia Appendix 3).

Weight Determination

A sample proforma was developed to obtain expert ratings for various subindicators, indicators, key indicators, and categories. This proforma can be self-administered via email or used for in-person interviews, thereby providing efficient data collection from experts. It also allows the experts to provide additional information such as variables, data sources, and case studies (Multimedia Appendix 4).

XSL•FO
RenderX

Procedure to Calculate Expert Weights

There are two mathematical weightage methods for calculating the indicator scores: (1) fuzzy extent analysis (FEA)–an efficient expert-based method and (2) Modified Entropy-Based Weightage Method (MEWM)–a data-driven method.

Fuzzy Extent Analysis (FEA)

The efficient expert-based FEA rating method used to calculate the indicator scores requires consultations with experts from diverse *One Health* sectors for ascertaining the priority or the weightage of different subindicators and indicators. The FEA is a multicriteria decision-making method that integrates both qualitative and quantitative approaches [9]. Data collection for expert-based rating can be conducted using the provided proforma. A pairwise comparison was used for developing the proforma, that allowed the experts to rate the metrics [10]. It is an effective way to gather opinions from many experienced experts, especially for complex decision-making problems involving multiple risks. The following is an example of how expert-based ratings are converted into weights. A pair of subindicators within an experts' domain will be provided to them for pairwise comparison using linguistic ratings, which are then translated into an existing numerical scale of relative importance [10]. Similarly, all possible combinations of subindicators will be provided to the expert for comparisons. The linguistic ratings obtained from an expert, such as equally important, moderately important, strongly important, and extremely important will be converted into numerical ratings. These numerical ratings are further converted into comparison matrix. Further, the consistency of the responses by the experts will be checked using the consistency ratio. After ascertaining the consistency of the responses, the next step will be "fuzzification" (ie, conversion of the numerical ratings to fuzzy numbers). To do this, we will use triangular fuzzy numbers (defined as a generalization of real numbers representing a set of possible values with weights, or membership functions) to calculate the weightage of different subindicators and indicators. Similarly, weights for the key indicators and categories may be obtained using this efficient expert-based weightage mechanism, or through a format similar to panel discussions which can be conducted to obtain the weights from the domain experts.

Steps involved for OHI calculation using FEA

Step 1

1. Consult with domain experts from diverse sectors relevant to *One Health*.
2. Based on the systematic implementation of the semistructured interviews and a modified Delphi method, pairwise comparisons are made on the importance of each pair of parameters.
3. Given a hierarchy with n parameters, there will be n(n – 1)/2 pairwise comparisons.
4. The comparisons are rated linguistically.
5. Linguistic variables are converted to numerical ratings using scale of relative importance [11].

Step 2

A comparison matrix will be established for each expert using the ratings from the proforma. The comparison matrix for the kthexpert is given by:

A~k=[aij] represented as:

A~K=(1 a~1n a~n1 1)

The reciprocal matrix is denoted by [A~K]-1.

The reciprocity properties are given by:

$$aij\ -1=1aij,\forall\ i,j=1,2,3,...n$$

Step 3

It crucial to maintain consistency for expert-based weightage process to derive the indicator scores. To achieve this, a Consistency Index (CI) was calculated to evaluate the consistency of the comparison matrix [10]:

$$CI=\lambda max-nn-1$$

where λmax is the maximum eigen value and *n* is the dimension of the comparison matrix [9].

The consistency ratio is given by:

$$CR=CIRI$$

where, RI is the random consistency index which depends on *n* [10].



If CR <0.1, the judgements are considered consistent [12].

Step 4

1. Convert the ratings in the comparison matrix to triangular fuzzy number.
2. A fuzzy number serves as a tool to express values that are uncertain or imprecise particularly in the context of fuzzy set theory.
3. Unlike a value, it accommodates varying degrees of belongingness enabling it to account for the ambiguity often found in practical scenarios.
4. Fuzzy sets can compensate for the inconsistency and imprecisions in human judgements rather than random or stochastic ones.
5. Fuzzy numbers are depicted by a set of possible values having their own membership function ranging from 0 to 1.
6. A triangular fuzzy number is represented by [floor value, average value, ceiling value] ie, (l,m,u) with the member function as:

$$\mu M(x)=\{xm-l-lm-l,x\ [l,m]xm-u-um-u,x\ [l,m]0,otherwise$$



[13].

Step 5

To calculate weights from the comparison matrix, FEA is applied [13].

XSL•FO
RenderX

In extent analysis, let X={x1,x2,...,xn} be an object set and G={g1,g2,...,gn} a goal set. For each object, extent analysis is performed for each goal gi.

The value of fuzzy synthetic extent is given by:

$$Si=\sum j=1mMgi\ j\otimes[\sum i=1n\sum j=1mMgi\ j]-1$$

where,

$$\sum j=1mMgi\ j=(\sum j=1mlj,\sum j=1mmj,\sum j=1muj)$$

We can calculate the ratings in the triangular fuzzy number format as shown above leading to a weight derived from these triangular weights.

A pairwise comparison of fuzzy weights needs to be performed and the computation of the degree of possibility of them being greater than the fuzzy weight will be obtained. The minimum of these possibilities is used as the overall score for each criterion *i*.

To compute the value of the fuzzy synthetic extent, Si for the ith object is as follows:

$$Si=\sum j=1mMi\ j\otimes[\sum i=1n\sum j=1mMi\ j]-1$$

where,

$$\sum i=1mMi\ j=(\sum j=1mli\ j,\sum j=1mmi\ j,\sum j=1mni\ j)$$

$$[\sum=1n\sum j=1mMij]-1=(1\sum=1n\sum j=1muij,1\sum=1n\sum j=1mmij,1\sum=1n\sum j=1mlij)$$

Step 6

The fuzzy synthetic extent constitutes three values which will be averaged out to acquire a single value for the weights from them.

Weights = $\sum i=1kWi\sim/k$

Step 7

Normalize each weight by dividing the individual weights by the sum of all weights.

$$\text{Normalized weight } i=Wi/\sum i=1kWi$$

Step 8

If the ratings have been collected from multiple experts for the same indicator, either of these methods can be used to obtain optimized ratings:

Method 1

1. Calculate individual normalized weights using Steps 1-7 for every expert rating
2. Average the normalized weights to obtain optimized ratings for the indicators

Method 2

- In continuation to Step 5, after calculating the fuzzy synthetic extent for every expert input individually,

1. Compute the degree of possibility of
   - S2(l2,m2,u2)≥S1(l1,m1,u1)
   - where degree of possibility between two fuzzy synthetic extents is defined as:
   - $V(S2\geq S1)=\sup y\leq x[\min(\mu S2(y),\mu S1(x))]$
   - $V(S2\geq S1)=hgt(S1\cap S2)=\mu S2(d)$
   - where, $\mu S2(d)=\{1,\ \text{if } m2\geq m1\ 0,\ \text{if } l1\geq u2l1-u2m2-u2, \text{otherwise}$
   - and› d is the ordinate of the highest intersection point d between μS1 and μS2.



1. Compute the degree of possibility for a convex fuzzy number to be >k convex fuzzy numbers Si(1, 2, …, k)
   - V[(S≥S1,S2,,,Sk)=V[(S≥S1)and(S≥S2)..and(S≥Sk)]=minV(S≥Si),i=1,2,,,k
2. Compute the vector W'.
   - W′=(d′(A1),d′(A2),...,d′(Ak))T
   - d′(Ai)=minV(Si ≥ Sj) Where
   - for i=1, 2, …, k and j=1, 2, …, k and i ≠ j
   - Normalized vector, W=(d(A1),d(A2),…,d(Ak))T
   - W is a nonfuzzy number calculated for each comparison matrix [14],
   - Now, the weights have been calculated through FEA, thus using these weights and an appropriate weight accumulation formula, OHI can be calculated.

Step 9

For the accumulation of the indicators and the weights, the following formula can be used:

$$Sij=\{0Xij-Xworst,jXbest,j-Xworst,j100*100$$

where Sij denotes the normalized score for the jth sub indicator of ith indicator; Xij denotes the original values for the jth sub indicator of ith indicator, Xbest,j denotes the best value for the jth sub indicator of the ith indicator, Xworst,j denotes the worst value for the jth sub indicator of the ithindicator. In cases where no data is available for the sub-indicators, substitute Sij with 0.

The weighted sum of the scores can be given by:

$$\text{Indicator Scoreih}=\sum 1hmhSijh*Wjh,\sum 1hmhWjh=1,$$

where, m depicts the number of subindicators under the hth indicator, jh depicts the jthsubindicator under the hth indicator, Sijh depicts the score of the jhth sub indicator under ith indicator; Wjh depicts the weight of jhth subindicator [8].

The procedure is repeated stage-wise across sub-indicators, indicators, key indicators, and categories to calculate the OHI.

Modified Entropy Based Weightage Method (MEWM)

This approach requires the indicators values for OHI calculation. Weightage mechanism is backed by the entropy-based weightage method for the indicator value and is a preferred method due to the ease of calculation and does not require expert-based rating.

Step 1

First, normalization (a systematic process of organizing data in a database to make it more flexible and cohesive) of the values of the sub indicators will be done using the following formula,

$$rij=rij=xij-xmin,jmax\ xj-minxj$$

where max xj and min xj are the maximum and minimum values among the alternatives for indicator *j* [15] .

XSL•FO
RenderX

Step 2

The entropy Eij of each subindicator *i* from the indicator *j*, the entropy Eij of each indicator was determined from the normalized values rij as formulated:

$$Eij = -\ fij * \ln fijln\ n \quad (i=1,\ldots,n,\ j=1,\ldots,m)$$

Where $fij = rij / \sum_{i=1}^{n} rij$

For the cases where fij= 0or is not available, the entropy becomes not defined and, in those cases, substitute the entropy with 0.

Step 3

For the applicability of the method, the weights wij are computed as defined below:

$$wij = 1 - Eijn - \sum_{i=1}^{n} Eij\ ,\ \sum_{i=1}^{n} wij = 1,\ (i=1,\ldots,n)$$

Step 4

The above formula can be used at every step to calculate the relevant weights,

$$Vij = wij * rij \quad (i=1,\ldots,n,\ j=1,\ldots,m)$$

where rij is the standardized value of the ith sub-indicator for the jth state and wij being the weight calculated for the ith sub-indicator for the jth state; vij is the weighted indicator value.

$$ri,\ j+1 = vj = \sum_{i=1}^{n} vij$$

where vij is the weighted indicator value for the ith subindicator for the jth state and vj is the index value of the indicator j using which, rj is calculated.

One can use this formula to arrive at the weighted value of the subindicators and then use those subindicator values to derive weights for the indicators and similarly repeating the process for all available data. These steps will transition stagewise in an agglomerative way from the scores of the sub indicators of a particular indicator to the OHI.

Ethical Considerations

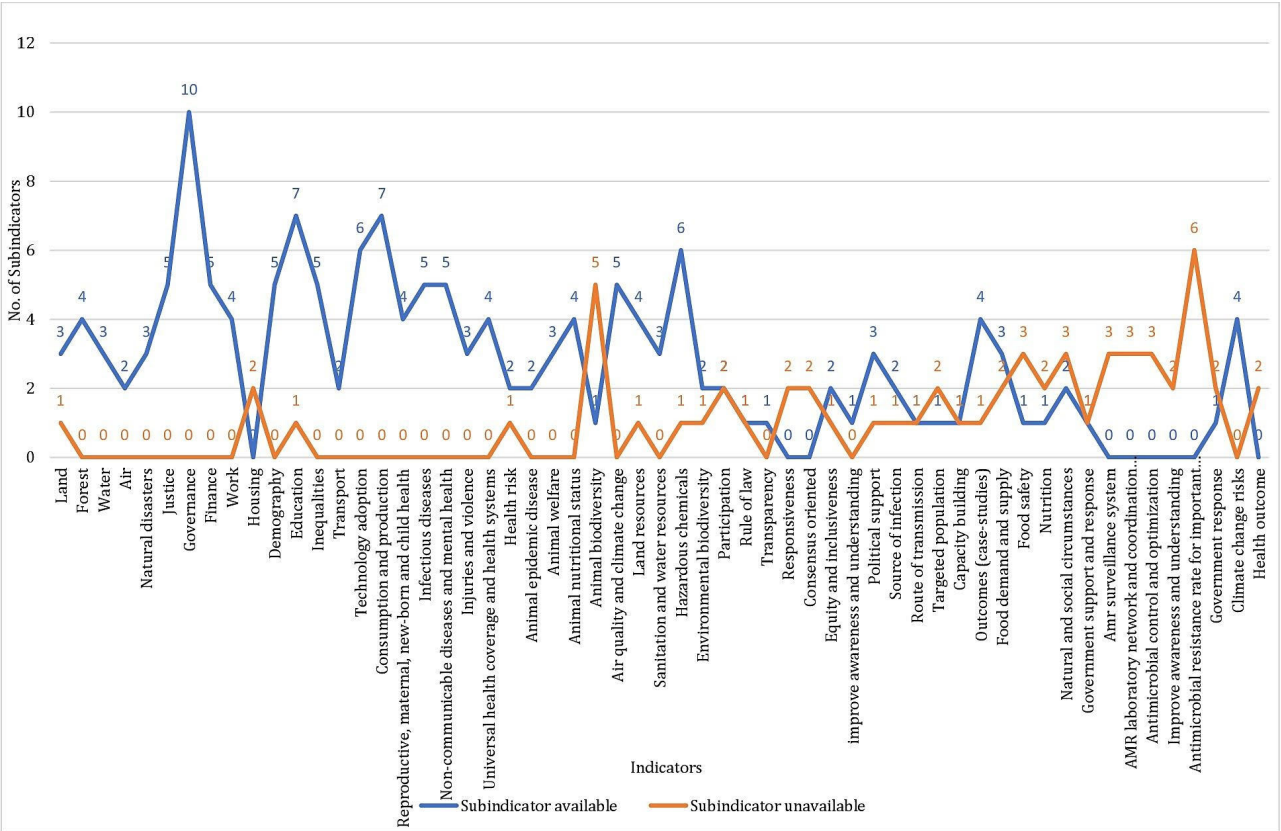
A formal ethical review was waived. This study was not submitted for review as the Tata Institute for Genetics and Society (TIGS) adheres to Indian Council for Medical Research ethical guidelines, which allow exemption for minimal-risk studies using publicly available data. Further, our manuscript is solely based on secondary data from publicly available databases and these were appropriately cited to maintain research integrity.

Results

Primary findings

Using secondary data sources, a country-specific dataset was developed for India. From this dataset, the indicator values for 156 of the 216 subindicators were gathered, while the remaining 60 subindicators lacked data (Figure 1 and Multimedia Appendix 5). This India-specific dataset reflects data gaps, that is, inconsistency in data availability and areas where there is absence of data requires planned interventions by governance systems. In some cases, due to the absence of historical data, the scores could not be computed.

Figure 1. Data availability across indicators. Blue represents the number of subindicators for which the data is available and orange represents the number of subindicators for which the data is unavailable.



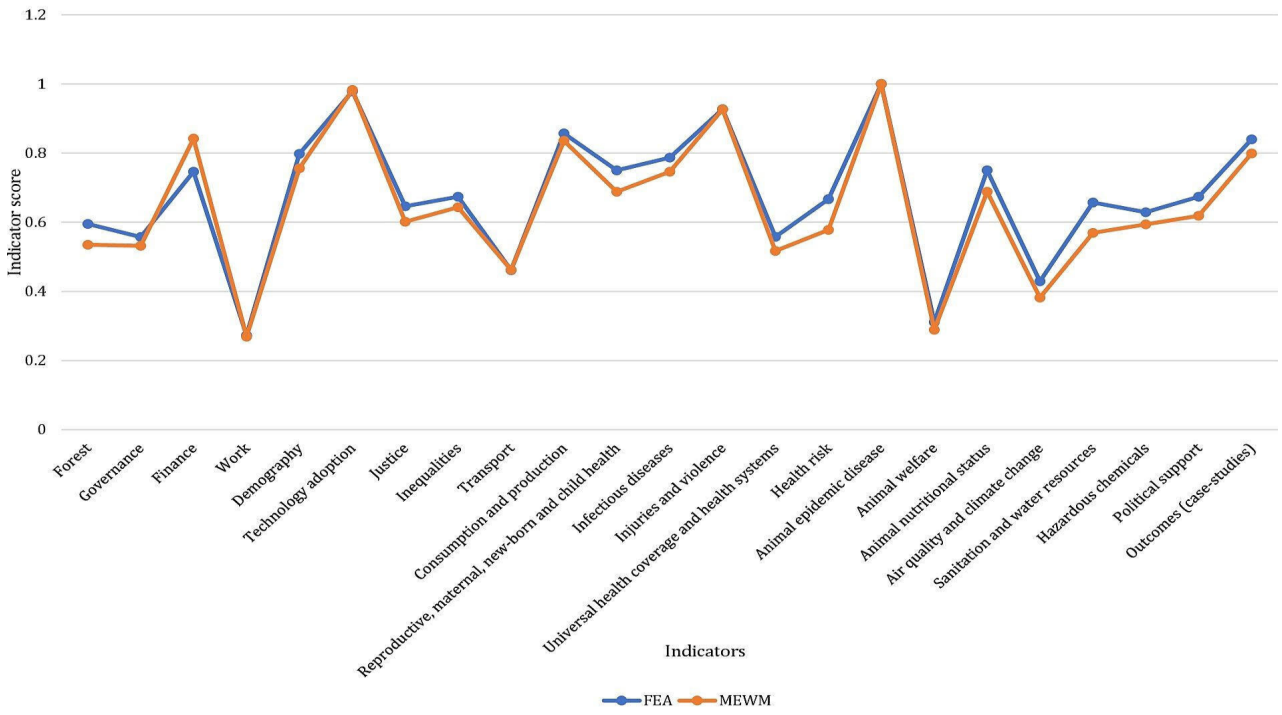
XSL•FO
RenderX

Pilot Analysis

Additionally, to evaluate the efficiency of weightage methods, a comparative analysis was conducted between FEA and MEWM. Scores were calculated for 23 indicators for which data for all or most sub-indicators were available (Figure 2). The findings demonstrate that the indicator scores obtained using the two weightage methods, FEA and MEWM for 23 indicators are significantly consistent and thereby render the two methods reliable. Further, to check the OHI calculator's applicability and utility in the Indian context, a correlation analysis was performed between 20 indicator scores and budget allocations for the years 2022 - 2023, 2023 - 2024, and 2024 - 2025 (Multimedia Appendix 3). The differences in budget allocations for each consecutive year were calculated, aiding in understanding the correlation between the indicator scores and the changes in budget allocations (Figure 3 and Figure 4). Table 1 shows that increases in budget allocations over consecutive years were associated with corresponding increase in indicator scores, and vice versa.

This emphasizes the importance of budgetary allocation in the performance of indicators, thereby impacting the overall performance of *One Health* in India. The OHI calculator has the potential to perform such nuanced correlation analysis; in this case, between budgetary allocation and performance of the indicators through their scores. This can serve as a valuable insight for policymakers and stakeholders alike in prioritizing sectoral interventions related to *One Health*.

Figure 2. Comparison of indicator scores using FEA and MEWM. FEA: fuzzy extent analysis; MEWM: modified entropy-based weightage method.



XSL•FO
RenderX

Figure 3. Budgetary differences for the years 2023 - 2024 and 2022 - 2023 against the indicator scores for consecutive years of budget allocations using the FEA and MEWM. A positive correlation is observed for both methods, indicating that an increase in budget allocation positively impacts the indicator scores. FEA: fuzzy extent analysis; MEWM: modified entropy-based weightage method.

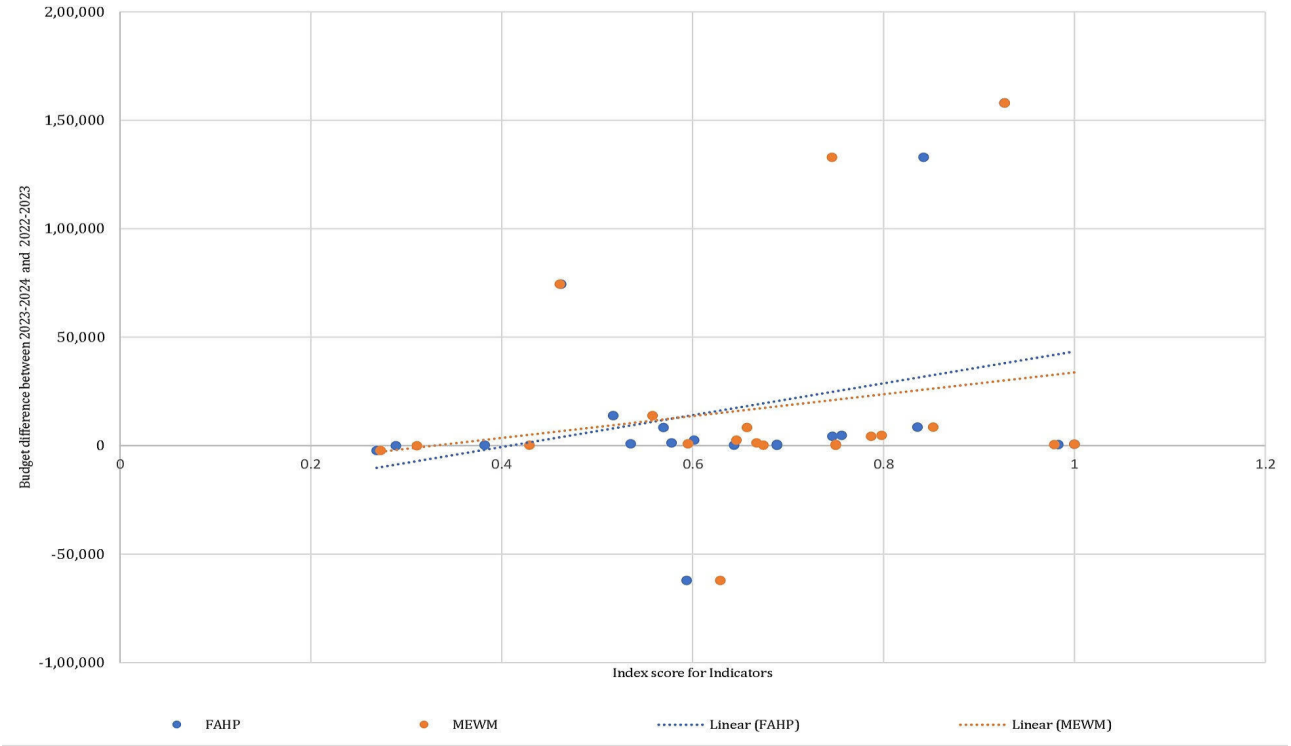
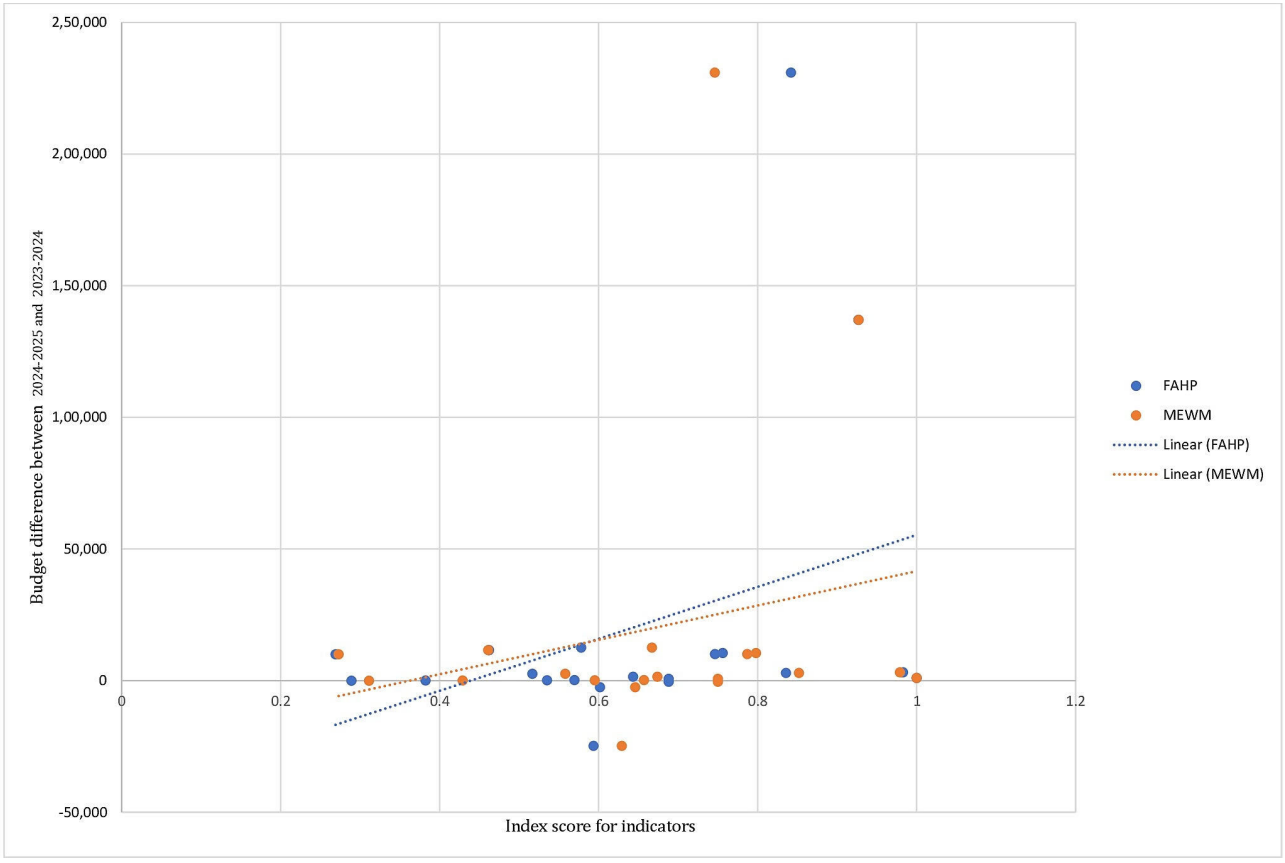


Figure 4. Budgetary differences for the years 2024 - 2025 and 2023 - 2024 against the indicator scores for consecutive years of budget allocations using the FEA and MEWM. A positive correlation is observed for both methods, indicating that an increase in budget allocation positively impacts the indicator scores. FAHP: fuzzy analytic hierarchy process; MEWM: modified entropy-based weightage method.



XSL•FO
RenderX

Table . Correlation between the increase in budget and indicator scores.

Difference between years	Indicator scores (FEA)[a]	Indicator scores (MEWM)[b]
2024 - 2025 and 2023 - 2024	0.353505892	0.22310977
2023 - 2024 and 2022 - 2023	0.311369994	0.20383558

[a]FEA: fuzzy extent analysis.

[b]MEWM: modified entropy-based weightage method.

Discussion

Summary

The OHI can be computed at various levels of governance in India, which is constituted of 28 states and 8 Union Territories. Its implementation depends on data availability, such as national, state, district or even block or village/panchayat levels. By calculating these values locally and then aggregating them, an accurate national value can be derived, which highlights demographic variations and provides a more precise measurement. For state-level calculations, population density can serve as a key weighting factor. Collaborative monitoring systems are now recognized as essential for effectively managing pandemics and outbreaks [3]. Thus, developing a OHI calculator can contribute to improved implementation of policies using a participatory *One Health* approach. However, the value of such an index extends beyond compiling data; it has the potential to revolutionize our understanding, management, and actions concerning the complex web of factors that influence global health. The process of calculating a OHI for India sheds light on critical areas requiring systematic interventions, aided by policy decisions, particularly regarding resource allocation and strengthening of governance systems. This framework also promotes a collaborative data federation model to address data gaps—such as incomplete data, lack of timely data, and the absence of appropriate data—and correspondingly advocates for data transparency and ethical data management. Implementing a collaborative data federation model and maintaining consistent data collection will help address these challenges and establish a historical dataset for indexing indicators. This approach encourages intersectoral communication and multistakeholder engagement, garners interest from governance systems, and builds momentum toward improving poorly performing indicators, thereby achieving better *One Health* outcomes for the country. A study focusing on mitigating zoonotic disease risk using the economic approach stated that allocating budget and resources strategically at a higher level secures sufficient funding to manage diseases along the livestock value chain, leading to improvements in human health [16]. As shown in Table 1, increasing the budget over consecutive years has led to higher indicator scores and vice versa, suggesting that budget increases for relevant ministries and departments contribute to improved indicator scores, thereby enhancing the OHI. Addressing the standardization of impact indicators and integrating new field knowledge are also essential.

In this study, we have provided a OHI calculator tool, demonstrating two weighting methods, FEA and MEWM, alongside the India-specific datasets that are accessible and efficient for use by multiple stakeholders such as government functionaries, policymakers, researchers, and institutes of disease surveillance and preparedness. While the scope of this study is to develop a OHI calculator for India, the objective is to provide a reliable, easy-to-use, and efficient tool that could be easily adapted by relevant sectoral stakeholders to compute empirical scores and plan informed and strategic interventions. We have also demonstrated OHI value calculations for India using both methods: the FEA method, which assigns equal weightage in the absence of expert weightage, yielded a score of 46.84, while the score obtained using the MEWM is 42.64. These values demonstrated that OHI calculation values fall within the range reported for the South Asia region, which is 35 - 50 as published in 2022, using the GOHI framework [8]. Further, it is also pertinent to conduct expert consultations to review the indicators adopted from the GOHI framework and to develop a list of indicators for India, which may be region-specific and locally adapted.

There are some limitations to be considered regarding the OHI calculator. For instance, the timeline for data availability is inconsistent across subindicators, and any gaps in data may impact the OHI score. Identifying reliable, validated data sources which is crucial for accuracy has been a challenge for many indicators. The selection of experts is a key factor, as their input determines the weightages assigned to different indicators. Any bias from the experts or the absence of a representative sample may result in a skewed OHI score. To avoid regional bias, it is crucial to ensure diverse representation from across the country for balanced input.

Conclusion

The OHI calculator identifies key areas that require concerted action from *One Health*-centric stakeholders. During its development, data gaps and deficiencies for crucial *One Health* indicators were identified, suggesting that data federation and open data access from governance systems and research organizations working in the public interest should emerge as actionable agendas for collaborative efforts. In this context, the methodology and framework for calculating a single OHI value require multisectoral experts to come together to improve the OHI for India by recognizing the need for sectoral data. This in turn, leads to identifying disparities, targeting interventions, monitoring health trends, and implementing other strategic efforts aligned toward the health equity paradigm. Additionally, the importance of budgetary allocation for improving indicator scores that contribute to OHI, serves as a reminder to policymakers about the value of empirical and evidence-based decision making. The process of calculating an empirical OHI value demands consistent public engagement and awareness of the interconnected nature of the *One Health* approach. Ultimately, this leads to better preparedness for handling future

XSL•FO
RenderX

pandemics, improved quality of life, and progress toward achieving the sustainable development goals.

Acknowledgments

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Global One Health Index (GOHI) framework list.
[XLSX File, 17 KB - ojphi_v17i1e65039_app1.xlsx]

Multimedia Appendix 2
India dataset.
[XLSX File, 208 KB - ojphi_v17i1e65039_app2.xlsx]

Multimedia Appendix 3
Budgetary allocation sheet vs. indicator scores.
[XLSX File, 13 KB - ojphi_v17i1e65039_app3.xlsx]

Multimedia Appendix 4
Sample proforma for expert rating.
[DOCX File, 55 KB - ojphi_v17i1e65039_app4.docx]

Multimedia Appendix 5
List of subindicators for which data is unavailable.
[XLSX File, 12 KB - ojphi_v17i1e65039_app5.xlsx]

References

1. One health. World Health Organization. 2017. URL: https://www.who.int/news-room/questions-and-answers/item/one-health [accessed 2024-07-19]
2. One world one health: history. Wildlife Conservation Society. URL: https://oneworldonehealth.wcs.org/About-Us/History.aspx [accessed 2024-07-19]
3. Mackenzie J, Jeggo M. The One Health Approach-Why is it so important? Trop Med Infect Dis 2019 May 31;2(4):88. [doi: 10.3390/tropicalmed4020088] [Medline: 31159338]
4. Horgan D, Hajduch M, Vrana M, et al. European health data space-an opportunity now to grasp the future of data-driven healthcare. Healthcare (Basel) 2022 Aug 26;10(9):1629. [doi: 10.3390/healthcare10091629] [Medline: 36141241]
5. Harris C, Green S, Ramsey W, Allen K, King R. Sustainability in health care by allocating resources effectively (SHARE) 1: introducing a series of papers reporting an investigation of disinvestment in a local healthcare setting. BMC Health Serv Res 2017 May 4;17(1):323. [doi: 10.1186/s12913-017-2210-7] [Medline: 28472962]
6. Clarke GM, Conti S, Wolters AT, Steventon A. Evaluating the impact of healthcare interventions using routine data. BMJ 2019 Jun 20;365:l2239. [doi: 10.1136/bmj.l2239] [Medline: 31221675]
7. High-quality health systems in the sustainable development goals era: time for a revolution - ScienceDirect. URL: https://www.sciencedirect.com/science/article/pii/S2214109X18303863?via%3Dihub [accessed 2024-07-19]
8. Zhang XX, Liu JS, Han LF, et al. Towards a global one health index: a potential assessment tool for one health performance. Infect Dis Poverty 2022 May 22;11(1):57. [doi: 10.1186/s40249-022-00979-9] [Medline: 35599310]
9. Peng G, Han L, Liu Z, Guo Y, Yan J, Jia X. An application of fuzzy analytic hierarchy process in risk evaluation model. Front Psychol 2021;12:715003. [doi: 10.3389/fpsyg.2021.715003] [Medline: 34630224]
10. Saaty TL. Decision-making with the AHP: Why is the principal eigenvector necessary. Eur J Oper Res 2003 Feb;145(1):85-91. [doi: 10.1016/S0377-2217(02)00227-8]
11. Zadeh LA. The concept of a linguistic variable and its application to approximate reasoning—I. Inf Sci (Ny) 1975;8(3):199-249. [doi: 10.1016/0020-0255(75)90036-5]

XSL•FO
RenderX

12. Vahidnia MH, Alesheikh A, Alimohammadi A, Bassiri A. Fuzzy analytical hierarchy process in GIS application. ISPRS Archives 2008;XXXVII(Part B2):593-596 [FREE Full text]
13. Chang DY. Applications of the extent analysis method on fuzzy AHP. Eur J Oper Res 1996 Dec;95(3):649-655. [doi: 10.1016/0377-2217(95)00300-2]
14. Ahmed F, Kilic K. Modification to fuzzy extent analysis method and its performance analysis. 2015 Presented at: 2015 International Conference on Industrial Engineering and Systems Management (IESM); Oct 21-23, 2015; Seville, Spain p. 435-438. [doi: 10.1109/IESM.2015.7380193]
15. Feng J, Guo Z, Ai L, et al. Establishment of an indicator framework for global One Health Intrinsic Drivers Index based on the grounded theory and fuzzy analytical hierarchy-entropy weight method. Infect Dis Poverty 2022 Dec 8;11(1):121. [doi: 10.1186/s40249-022-01042-3] [Medline: 36482389]
16. Häsler B, Gilbert W, Jones BA, Pfeiffer DU, Rushton J, Otte MJ. The economic value of One Health in relation to the mitigation of zoonotic disease risks. Curr Top Microbiol Immunol 2013;365:127-151. [doi: 10.1007/82_2012_239] [Medline: 24264885]

Abbreviations:

AMR: Antimicrobial resistance
FEA: fuzzy extent analysis
GOHI: Global One Health index
MEWM: Modified Entropy-based Weightage method
OHI: One Health Index

Edited by E Mensah; submitted 02.08.24; peer-reviewed by D John, XN Zhou; revised version received 16.12.24; accepted 19.04.25; published 25.06.25.

Please cite as:
Meganathan S, Katiyar A, Srivastava E, Mishra RK
One Health Index Calculator for India Using Empirical Methods for Policy Stewardship: Development and Usability Study
Online J Public Health Inform 2025;17:e65039
URL: https://ojphi.jmir.org/2025/1/e65039
doi:10.2196/65039

© Saveetha Meganathan, Arpit Katiyar, Esha Srivastava, Rakesh Kumar Mishra. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 25.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Issues in Identifying Strategies for Youth Mental Well-Being in Stockholm Municipalities Using Participatory Sessions and Text Mining: Qualitative Study

Harsha Krishna, MTech; Adam S Darwich, PhD; Sebastiaan Meijer, PhD

Division of Health Informatics and Logistics, KTH Royal Institute of Technology, Hälsovägen 11C, Huddinge, Stockholm, Sweden

Corresponding Author:
Harsha Krishna, MTech
Division of Health Informatics and Logistics, KTH Royal Institute of Technology, Hälsovägen 11C, Huddinge, Stockholm, Sweden

Abstract

Background: Socioeconomic and environmental factors influence youth mental well-being. Promoting mental well-being is essential to support youths' development toward adulthood with good mental health. Different Stockholm municipalities have adopted strategies to promote youth well-being. However, contextualizing and perceiving goals and mechanisms at the local municipal level is difficult. Thus, comparing or tracking their conception, purpose, and characteristics has been challenging.

Objective: We aimed to use data visualizations developed from a fusion of data sources to facilitate stakeholder conversations on promoting youth mental well-being within a municipality. We strive to demonstrate our methodology of using data visualizations as "boundary objects," which are cognitive artifacts that bridge knowledge from various domains to elicit understanding from specialized and siloed parts of a health delivery system.

Methods: Stakeholders from the municipalities of Lidingö and Nynäshamn participated in the study. A total of 15 workshops were conducted: 6 with only Lidingö participants, 6 with only Nynäshamn participants, and 3 with mixed participants. The sessions were conducted via Microsoft Teams or as physical sessions in Swedish and lasted between 60 and 90 minutes. Interactions were recorded with consent from participants. Recordings were transcribed using Amberscript software. We used matrix factorization with Kullback–Leibler divergence to extract 1000 features and created 10 topic clusters with 20 top words. We used the identified words and phrases to backtrack within the transcripts and to identify dialogues where they were used. We summarized participants' interactions across all the workshops to identify factors or strategies discussed for youth well-being.

Results: Participants noted that these sessions allowed them to contextualize their local observations from municipalities relative to the status of other municipalities in the national statistics. They indicated that they conceptualized well-being differently in their respective municipalities and between different professional backgrounds, and the sources of stress for youth differed. They noted the differences in the strategy and data collected for tracking youth well-being. Promotion of sports was a common strategy, while options for leisure activities differed between municipalities and professions.

Conclusions: Based on our observations and analysis of the transcripts from participatory workshops, we observed that the data-driven visualizations helped stakeholders from different departments of Lidingö and Nynäshamn municipalities to identify and bridge knowledge gaps caused by data silos. Participants noted proposals to modify future surveys and identified that this approach to visualizations would help them to share knowledge and maintain a long-term and sustainable collaboration across departments.

(Online J Public Health Inform 2025;17:e66377) doi:10.2196/66377

KEYWORDS

youth; mental wellbeing; Stockholm; participatory workshop; data driven; visualizations; boundary objects; text mining; municipalities

Introduction

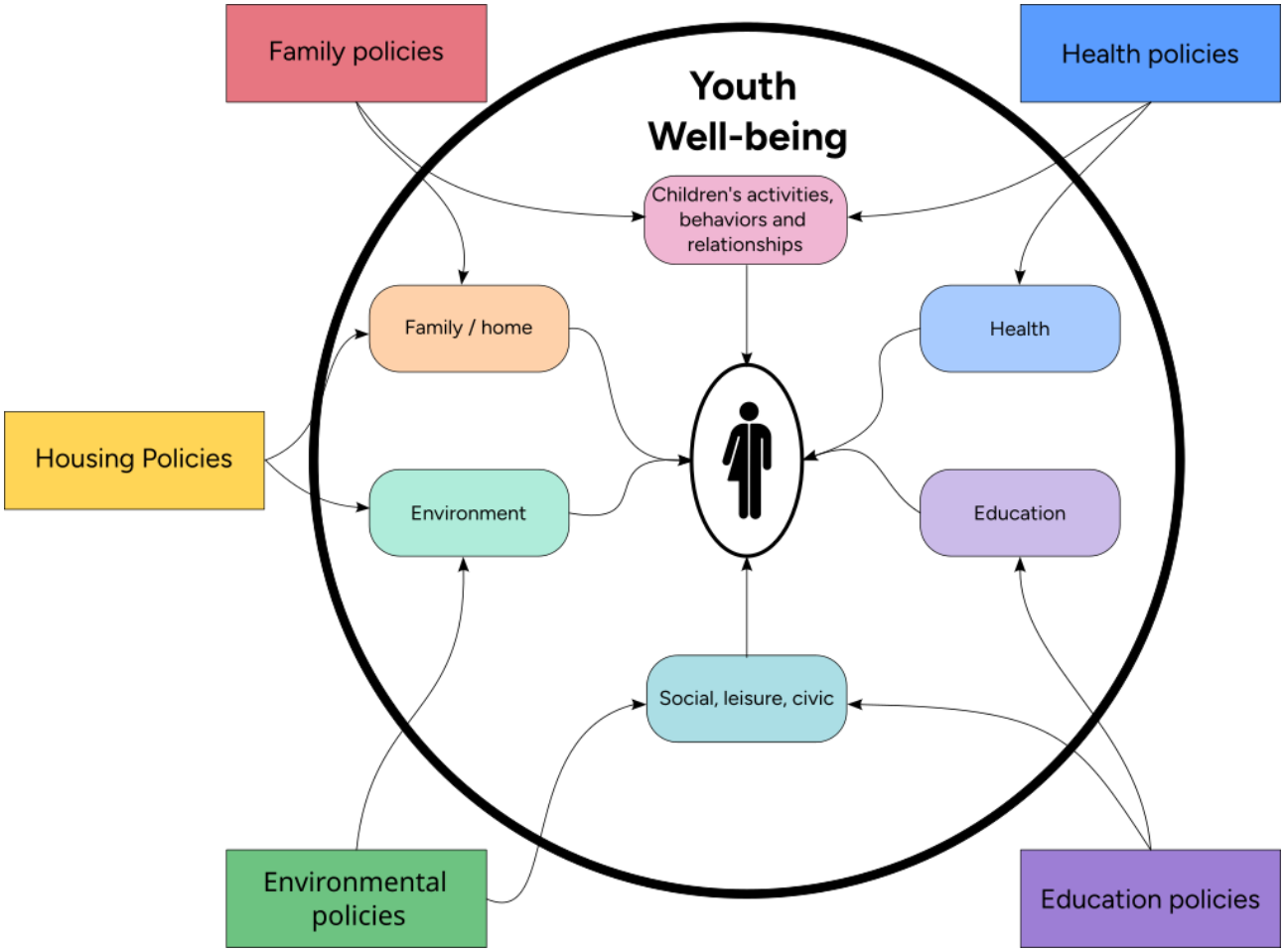
Sweden follows the World Health Organization (WHO) and Organisation of Economic Co-operation and Development (OECD) definition of good mental well-being as "a state of well-being in which the individual realizes his or her abilities, can cope with the normal stresses of life, can work productively and fruitfully, and can make a contribution to his or her community" [1-3]. The definition uses the 2-continuum model identified by Keyes et al [4], which states that individuals can be classified based on their mental illness and mental health status. According to this classification, a person with high mental health and low mental illness is stated to be "flourishing," and the contrary is said to be languishing. A person with high mental

XSL•FO
RenderX

health but enduring mental illness is considered flourishing with mental illness. It demonstrates that there are different factors, both genetic and environmental, that contribute to an individual's well-being. It also argues that despite enduring mental illness, an individual can lead a meaningful life and flourish. Targeting individuals when they are young ensures a preventative approach to well-being in the future by equipping them with strategies to manage and seek help early. Based on this goal, the Region of Stockholm adopted a goal to improve the mental well-being of its youth population.

Figure 1 lists some factors that potentially affect an individual's well-being, including environmental and physiological aspects, such as access to education, leisure time, and so on. These factors are influenced by various programs and policies implemented and monitored by local municipalities, as described in previous studies [2-8]. The work done by Moustaid et al [6] modeled the factors, such as education, social environments, schools, labor markets, and health care provisions, that affect mental health in cities. They demonstrated that poor mental health in cities could result from poor coordination between different city infrastructures. This was explored using a systems dynamics–based model in a participatory approach with other stakeholders.

Figure 1. Parameters involved in determining youth well-being, adapted from the Organisation for Economic Co-operation and Development [2].

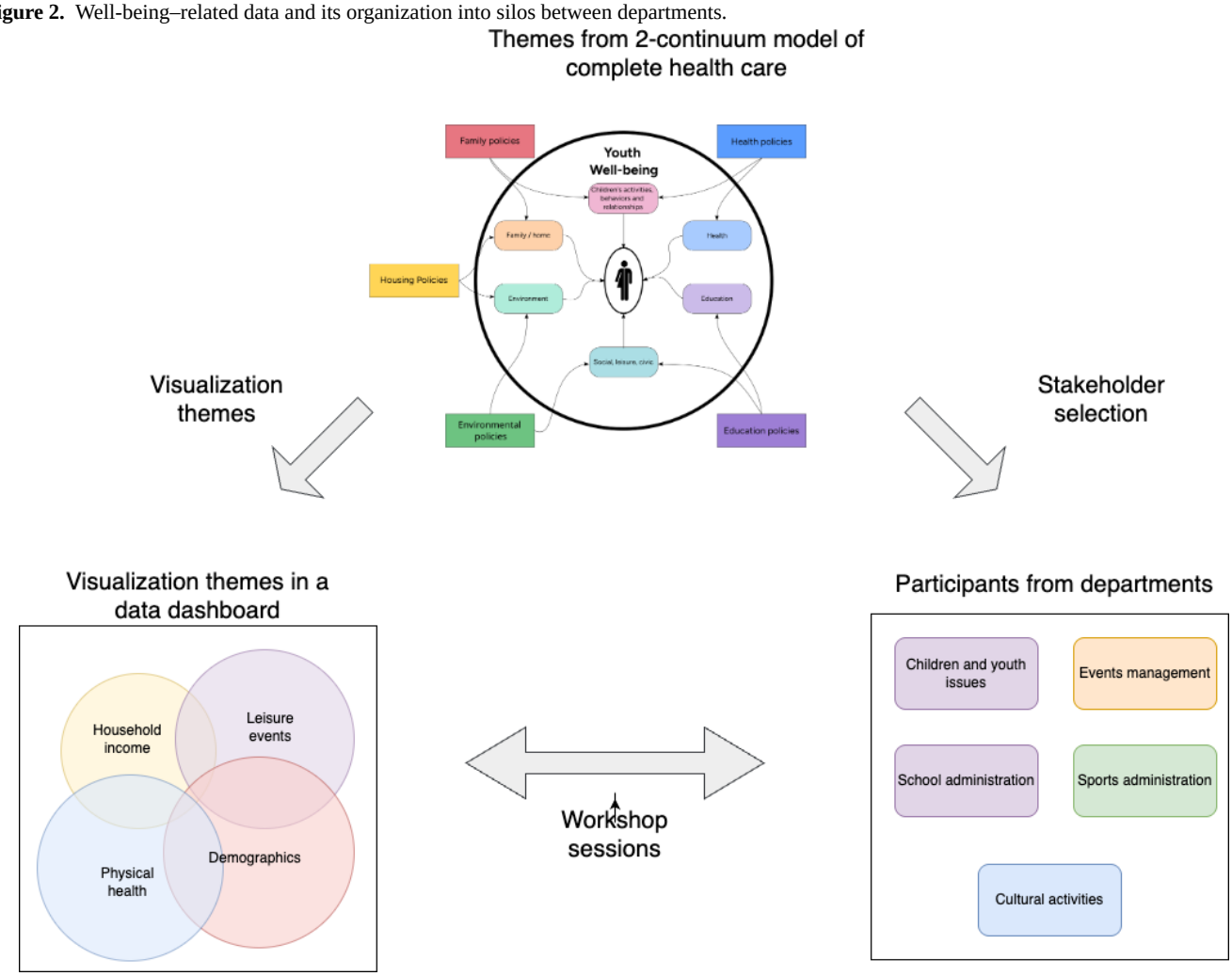


Typically, different departments in a municipality would be responsible for monitoring or implementing domain-specific strategies. This has led to data silos and myopic perspectives within departments in the municipal system. Well-being promotion strategies tend to cut across various departmental responsibilities. Therefore, a participatory session would bring this siloed data together and make it understandable for all domains.

As Figure 2 shows, no shared data infrastructure monitors all the identified dimensions. We used a fused dataset, which combined all the data available in a network database, to query across them.

XSL•FO
RenderX

Figure 2. Well-being–related data and its organization into silos between departments.



To help understand the goals and perspectives at the municipal level, we propose to use visualizations generated from datasets that influence mental health as "boundary objects." Boundary objects are defined as objects that live in multiple social worlds, each with a different identity. For example, an engineering drawing acts as a boundary object in manufacturing processes. It allows designers and fabricators to discuss challenges in their departments while working toward the same objective. Boundary objects and the theories of knowledge boundaries have been used to help work with complex interconnected systems, as described in previous studies [9-12], to further demonstrate and develop the theory.

Boundary objects help contextualize knowledge when working with expert stakeholders from different domains. It was proposed in the seminal paper [10]. The authors describe boundary objects as plastic enough to adapt to the local needs and constraints of several parties using them, yet robust enough to maintain a common identity across sites. Thus, they satisfy the informational requirements of various communities while still having distinct meanings in each of them. Boundary objects emerge through work processes. Boundaries refer to the overlap of areas between different communities.

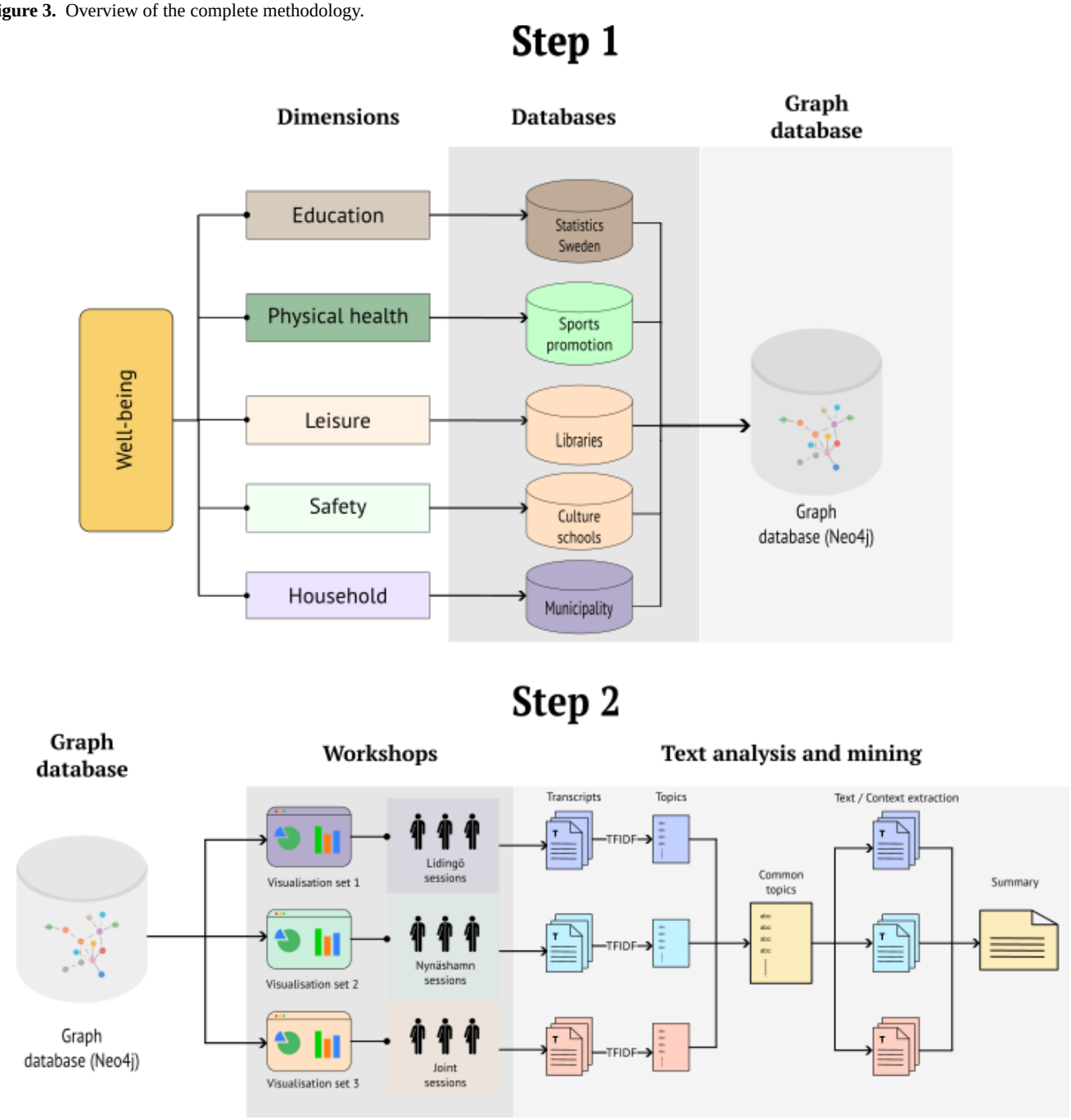
We propose using boundary objects to collaborate with different municipal departments. Data-driven visualizations can be used as boundary objects in participatory sessions, enabling different experts to discuss youth mental well-being. Authors in [13-15] have used visualizations as boundaries for building models in group sessions. Figure 3 illustrates an approach for using visualizations in a workshop. In the first step, a fused dataset is developed using data from national statistics and data from different municipal departments.

In the second step, various visualizations are generated based on different themes. The visualizations can be used within a workshop to discuss participants' experiences concerning various well-being promotion programs.

We built a fused dataset from separate sources to explore the different promotion programs and their reasons. We created it using a network database, Neo4j [16], which we explained in detail in our previous work. We used the fused dataset to build cross-sectoral visualizations, which we used in workshop sessions with participants from 2 municipalities, including the youth.

XSL•FO
RenderX

Figure 3. Overview of the complete methodology.



Methods

Overview

Figure 4 illustrates the research method. Two municipalities, Lidingö and Nynäshamn, agreed to collaborate on the project after a general request to all the municipalities of Region Stockholm. We collected data from national statistics [17] and different departments of Lidingö and Nynäshamn. We also received data from social services, schools, and other surveys run by the municipalities with their youth population.

For our study, we studied the population aged between 16 and 29 years. This was based on initial research and discussions with the municipalities. We visualized data from 2018-2019 and 2022-2023 to demonstrate changes observed in the data collected from youth between 19 and 29 years, split by sex. We note that the data from 2020 and 2021 would have recorded the impact of the global COVID-19 pandemic. While Sweden did not implement a complete lockdown, education was shifted to an online format, which could have caused variations in mental health during that period.

For each municipality, we developed a set of different visualizations based on the factors identified for youth well-being, as shown in Figure 1, and based on the various participants for a given workshop session.

We created workshop assets, which included (1) an introductory presentation, (2) a set of data visualizations, (3) a video summary of the data being presented and access to the data dashboard, and (4) a set of discussion topics for the workshop.

XSL•FO
RenderX

Figure 4. Research methodology.

XSL•FO
RenderX

A fused database allowed us to query data across different themes. For example, we could retrieve data on the education, environment, home condition, and neighborhood safety of a given age group of youth in a municipality from the same database.

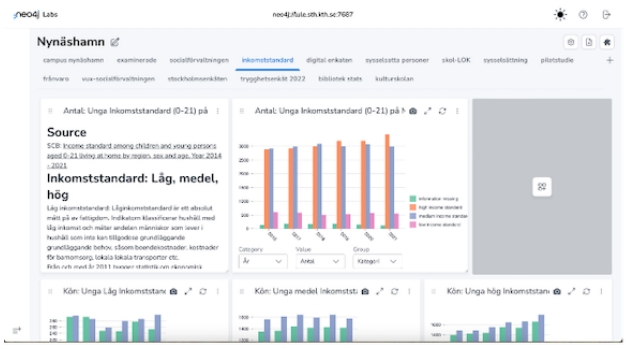
The visualizations included bar graphs, pie charts, and line graphs. We provided data sources with periods, local or national, and the name of the study, along with the visualizations. Multiple visualizations were arranged on different pages based on education, leisure activities, sports, and so on.

As per the definition of a boundary object, each dashboard was created with visualizations of data from national statistics and data from the participants' departments. The selections were based on (1) the municipality that participants belong to and participants' expertise who are attending the workshop.

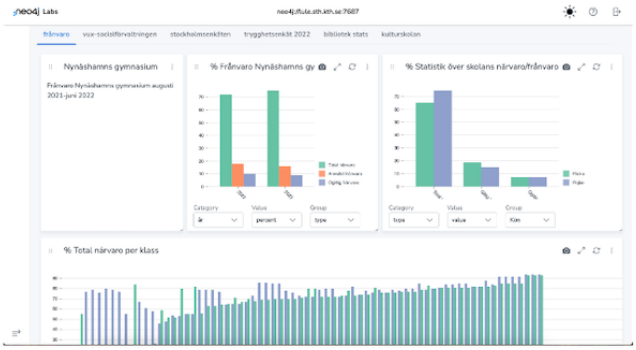
For example, for a session with Lidingö, if the participants were from sports organizations and clubs along with the municipal administrators and educators, a dashboard was constructed by combining data from sports organizations about youth between the ages of 15 - 25 years for the selected time period of 2019 - 2023 in Lidingö. Data on school attendance and academic performance, national statistics, and demographics of the youth in Lidingö would also be added and visualized. If any further programs or campaigns were organized, their performance was also added to the dashboard. This presented an overview of the leisure, education, and background of the youth population from similarly grouped populations.

Figure 5 illustrates an example of data visualizations developed on the Neo Dash platform (Neo4j Labs) [18] with the Neo4j graph database.

Figure 5. Dashboard illustrating income standard and high school attendance of youth in Nynäshamn municipality.



A. Dashboard illustrating income standard of youth in Nynäshamn municipality.



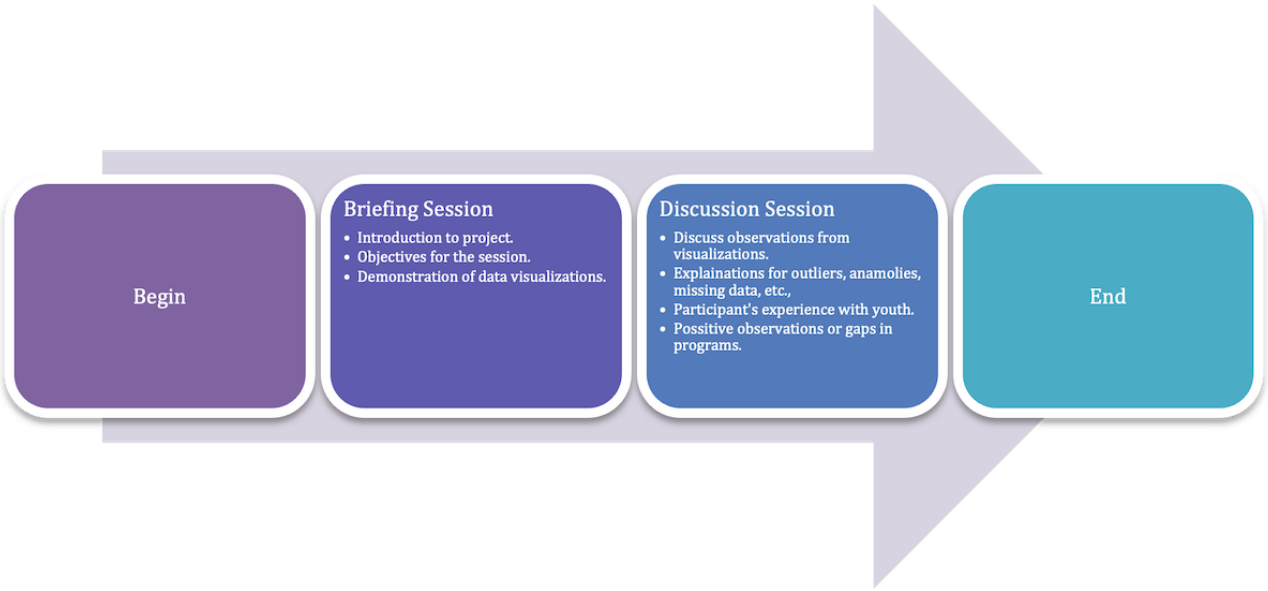
B. Dashboard illustrating attendance at high schools in Nynäshamn.

It demonstrates the income standard visualization for Nynäshamn municipality, which is targeted at policy makers and school personnel sessions. This would be combined with the visualization of school attendance and the activities arranged in Nynäshamn by the libraries, which target the youth. This allowed us to change and compose different pages and visualization sets based on the participants' backgrounds for a workshop. It also allowed us to construct new visualizations on demand during a workshop.

We began each workshop with a briefing followed by a discussion. Figure 6 illustrates the workshop's progress and various steps. The workshops were scheduled between 60 and 90 minutes. A screenshot from one such session is included in the Multimedia Appendices (Figure S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 6. Conducting a workshop session.



All workshops were conducted via Microsoft Teams except the first one with Lidingö, which was held as a physical session. The appendices list all workshops conducted, their themes, and participation details (Table S2 in Multimedia Appendix 2).

Participants' consent for recording was obtained for both online and physical sessions. Notes and audio were collected during all sessions. Online sessions were recorded using conference software and transcribed using the Amberscript software [19]. All the content and analysis were in Swedish. We have included the content in both Swedish and English in the Multimedia Appendices.

We used the minibatch nonnegative matrix factorization with the Kullback–Leibler divergence [20,21] to analyze the content and identify the most critical topics. We varied the n-grams, that is, the length of phrases, between 2 and 3 words to identify the top words and phrases used in each text corpus under consideration. We developed Python scripts with scikit-learn libraries [22-25]. The text was processed by removing the timestamps from the transcripts and creating a stop-word list to avoid identifying pronouns, numbers, and member names. A list of all the Swedish stop-words identified for processing is provided in the Multimedia Appendices (Table S3 in Multimedia Appendix 3).

We set the system to sample the corpus for 2000 samples and extract at least 1000 features. We then extracted at least 10 topics or clusters with 20 top words in batches of 128. The algorithm extracted the top 20 clustered words and formed 10 such clusters. Table S1 in Multimedia Appendix 4 lists the top words identified in Swedish. We traced the identified words back to the transcripts to identify and summarize the context of the discussions (Multimedia Appendices 5 and 6). We summarize the discussions as follows:

1. How participants approached defining well-being.
2. Tracking youth well-being with currently available data.
3. How leisure time is used by youth.
4. Current strategies to improve youth well-being.

Ethical Considerations

Verbal consent was obtained from all participants at the recording of each workshop session for analysis and research purposes. No identifiable information has been used in the analysis apart from the target municipality for the sessions. Participants could leave any of the sessions at any time before the recording began. All analysis is conducted without specific names. The data used for visualizations are from publicly available sources from Statistics Sweden.

Results

The top 10 topics from the corpus with the Lidingö municipality workshop sessions are illustrated in Figure S2 in Multimedia Appendix 7. Similarly, the topics selected from the corpus of Nynäshamn and the joint workshop sessions are illustrated in Figures S3 and S4 in Multimedia Appendix 3 and Multimedia Appendices 8 and 9, respectively. The graphs represent the numerical score of the words as a combination of 2 to 3 words per the n-grams setting. Each cluster represents a topic cluster internally arranged according to the calculated score. The score is calculated using the ratio of term frequency, the number of times the term occurs in a document, and inverse document frequency, the inverse number of documents in which the term appears. A matrix is constructed for all such features identified in all the documents, which is then factorized to collect the topics. We used the identified words and searched the transcripts to identify and summarize the context of the discussions.

Youth from Lidingö expressed that a bad day would not define their mental health. They considered the persistence of a bad mood as poor mental health. They also indicated that individuals create a facade or hesitate to seek help. Routines were deemed essential to help with anxiety. It was difficult for them to differentiate between mental health and well-being.

Highly educated parents and household locations in Stockholm positively impact youth well-being. A recent study observed a similar trend [26]. Parental expectations of youth contribute negatively to youth well-being. In Sweden, the ninth-grade

XSL•FO
RenderX

scores determine admissions to high schools and are considered essential.

Personnel from both municipalities stated that attending school is the most protective factor for youth. Therefore, many efforts focus on keeping youth in schools, reducing absenteeism, and returning youth to receive an education. They discussed the inclusion of individuals with disabilities, families with members who have mental illness, and teenage parents.

Safety issues were not discussed or observed in either municipality.

Expert participants observed that although much data have been collected, converting it into knowledge has been difficult. Many were keen to use data to inform their work, but have found it hard to find tools to collaborate across departments. Quantitative data were rarely enough to generate insights. They discussed the importance of events involving different departments in performing cross-sectoral analysis.

Expert participants also noted that negative aspects of mental health had been historically tracked, but positive effects of well-being promotion were not included in current surveys. Youths aged between 21 and 29 years were not adequately tracked to observe their transition into adulthood. The Stockholm survey, conducted nationally every 2 years with adolescents, was an example of a well-defined survey constructed from consultations across departments. Participants indicated that parameters such as free-time activities should be tracked regularly within surveys, for example, the Stockholm survey.

Experts from both municipalities identified that different departments track the youths' free time. They indicated that while sports and outdoor activities improve youth well-being, parents sometimes need to participate in such activities. Such support could include transportation, organization, acquiring resources, and actively playing with them. This also allows parents to build relationships and trust across generations. Thus, free time available for parents also plays a role in the youths' ability to access such programs. Nynäshamn personnel observed that attending organized sporting events becomes increasingly stressful as youth age. The stress of balancing studies and planning a career takes precedence over leisure activities.

The youth of Lidingö had more options to spend their free time than the youth in Nynäshamn. Girls suggested the development of leisure parks. They preferred more social activities compared to boys. The female participants indicated they would like to gather and do activities within leisure parks, expressing that they tend to participate less in organized sports. Participants noted that developing leisure parks for mixed use by both older adults and youth to engage in leisure activities would help multiple individuals within a municipality. However, football was popular among both boys and girls.

Expert participants of both municipalities identified that maintaining long-term collaborations across departments is a challenge. Nynäshamn created an interdepartmental group, such as collaboration between the children and educational administration and business-labor market administration, to improve labor access for their youth. Youths aged 19 - 29 years who did not have a linear education path were provided support. A coordination agency that can direct various programs in Campus Nynäshamn has helped young people return to education. The program included financial assistance and education to enter the labor market. Programs such as Erasmus volunteer programs with visits to other countries have helped youth engage with studies and activities. Lidingö personnel also identified that summer jobs positively affected youth well-being. They allowed youth to gain professional experience and interact with people outside their social cliques.

Discussion

Principal Findings

In this work, we used data visualization as a collection of boundary objects in a structured workshop among participants to identify and understand the impact of programs on improving youth well-being in Lidingö and Nynäshamn municipalities of Sweden. We analyzed participatory workshop transcripts to identify the most critical topics discussed in all sessions. This enabled us to backtrack within the transcripts and identify the context for discussing the topic. We used the terms extracted to identify the sentences and contexts from the transcripts (Multimedia Appendix 7).

Visualizations provided the starting point for such discussions in each workshop. Almost all the participants observed that the visualizations helped them better situate their local datasets within the national statistics. Participants requested further fine-grained resolution for data visualizations and more combinations of visuals across departments for discussions. Participants in multiple sessions remarked that they could think of new ideas during these sessions.

The visualizations indeed acted as boundary objects. They allowed participants from different departments to trace individuals' paths while pursuing their education and the social support they may access as they grow. They also allowed for a conversation on the different contexts in which an individual is situated, such as individuals who need social support or motivation to continue higher education.

Due to the existing administrative structure, personnel tend to focus on their local departments and goals. Any collected data and the eventual knowledge derived from the same end up in data silos within departments. This causes a myopic view of the municipal system. Our data visualizations included data from allied parts of the municipality. This gave participants insight into how their department is situated in the overall system and the work done by allied departments. They were also able to identify modifications to national surveys to track the transition from teenagers to adults better.

We saw evidence for "a boundary object effect" as participants could identify missing information from current data infrastructure and surveys, such as the ability to track individuals who turn 18 years of age and, thus, exit the youth programs. They may still need support after they achieve legal adulthood. This is currently not possible directly due to differences in data infrastructure and departments. We noted that this identification was possible due to visualizations connecting data from one department. Visualizing data from within a given department

XSL•FO
RenderX

would not provide new insights for personnel from the same department.

Limitations

Our current work used 15 workshops with participants from 2 municipalities. Increasing the number of sessions or extending dialogues would produce more extensive transcripts. Thus, the size and quality of the discussions were limitations. A trained facilitator can help produce better-quality dialogues.

We created visualizations based on themes surrounding youth mental well-being. During the workshops, some participants requested more on-the-fly visualizations. While this was achieved, a better tool that can quickly adapt to requests without technical complexity would allow for improved interactions between data and participants.

For the text analysis, we relied on the Natural Language Toolkit (NLTK) library's stop-words for Swedish and our own set of identified Swedish stop-words. A better listing of stop-words would identify topics, avoiding common phrases (Multimedia Appendix 5).

Conclusions

We observed that using fused data allowed us to create visualizations from a given department and data about related system parts that promote well-being. The visualizations allowed participants from different departments to relate to perspectives from associated departments, demonstrating that such contextualized visualizations could perform the function of a boundary object. We observed this in the interaction involving the sports department in the Nynäshamn municipality when participants identified that parents' choice of sports and participation was necessary.

Both municipalities were able to contextualize well-being based on their local population demographics. Lidingö identified that the nature of stress the students feel relates to their academic performance and future aspirations. Nynäshamn identified that their youth need further motivation to continue their education and even aspire for a higher education. We observe here that although both municipalities have the same goal for achieving youth well-being when contextualized to their local situation, they present themselves as different challenges to the municipalities.

Participants from both municipalities (Lidingö and Nynäshamn) faced challenges with their youth populations, affecting their approach to promoting youth well-being. We observed that participants in both sessions requested new combinations for visualizations and dashboards, demonstrating variations in how they intend to use their data and knowledge.

Finally, both municipalities noted a need for changes to national surveys and data infrastructure collecting around the age group of 19 - 25 years. Experts from educational departments noted that at this transition, their needs and support would stop being tracked by youth-based institutions. They will be considered adults and must approach adult (or general public) institutions for any support. This prevents experts from a better handoff to the adult system.

We plan to develop tools that enable municipal personnel to create visualizations on demand during sessions, directly using large language models and fused network databases. This would enhance interaction levels and promote interdepartmental collaboration. With further sessions with other Stockholm municipalities, we hope to see further nuances related to local effects.

Acknowledgments

This project was a collaboration with Petra Lindfors and Kari Trost of Stockholm University, Elisabet Rendahl from Region Stockholm, and Jenny Furstenbach from Nyfikenhet AB. The authors wish to thank the many collaborators from the municipalities of Lidingö and Nynäshamn, Sweden. Region Stockholm funded this work within the Research and Innovation portfolio.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demonstrating visualizations of free time at Nynäshamn during a briefing session of a workshop.
[PNG File, 410 KB - ojphi_v17i1e66377_app1.png]

Multimedia Appendix 2
List of all workshops conducted with participants and theme.
[DOCX File, 15 KB - ojphi_v17i1e66377_app2.docx]

Multimedia Appendix 3
Swedish stop-words used for text analysis in addition to the Natural Language Toolkit.
[DOCX File, 15 KB - ojphi_v17i1e66377_app3.docx]

Multimedia Appendix 4
Top words identified from intersections of all topics and workshops (Swedish).

XSL•FO
RenderX

[DOCX File, 19 KB - ojphi_v17i1e66377_app4.docx]

Multimedia Appendix 5
Mapping for column names and file names for the matrix of occurrences of common topics extracted in all files.
[DOCX File, 13 KB - ojphi_v17i1e66377_app5.docx]

Multimedia Appendix 6
List of common topics extracted from all the files and their occurrences in the lines of files.
[DOCX File, 35 KB - ojphi_v17i1e66377_app6.docx]

Multimedia Appendix 7
Topic clusters identified from words in the transcripts of workshops with Lidingö.
[PNG File, 471 KB - ojphi_v17i1e66377_app7.png]

Multimedia Appendix 8
Topic clusters identified from words in the transcripts of workshops with Nynäshamn.
[PNG File, 492 KB - ojphi_v17i1e66377_app8.png]

Multimedia Appendix 9
Topic clusters identified from words in the transcripts of workshops with both municipalities.
[PNG File, 490 KB - ojphi_v17i1e66377_app9.png]

References

1. Begrepp inom området psykisk hälsa version 2020. SKR, Folkhälsomyndigheten,Socialstyrelsen. URL: https://skr.se/download/18.1c90461a17cda637a4363ded/1636721711698/PM_Begrepp-inom-omradet-psykisk-halsa-TA.pdf [accessed 2025-06-16]
2. OECD. Measuring What Matters for Child Well-Being and Policies: OECD Publishing; 2021. [doi: 10.1787/e82fded1-en]
3. García-Carrión R, Villarejo-Carballido B, Villardón-Gallego L. Children and adolescents mental health: a systematic review of interaction-based interventions in schools and communities. Front Psychol 2019;10:918. [doi: 10.3389/fpsyg.2019.00918] [Medline: 31068881]
4. Keyes CLM. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222. [doi: 10.2307/3090197] [Medline: 12096700]
5. Huppert FA. Psychological well‐being: evidence regarding its causes and consequences. Appl Psychol Health Well Being 2009 Jul;1(2):137-164 [FREE Full text] [doi: 10.1111/j.1758-0854.2009.01008.x]
6. Moustaid E, Kornevs M, Lindencrona F, Meijer S. A system of systems of mental health in cities, digging deep into the origins of complexity. Adm Policy Ment Health 2020 Nov;47(6):961-971. [doi: 10.1007/s10488-020-01035-0] [Medline: 32222843]
7. Lindfors P. Physiological correlates of mental well-being. In: Mental Well-Being: Springer Netherlands; 2013:259-274. [doi: 10.1007/978-94-007-5195-8_12]
8. Muwonge JJ, Dalman C, Burström B, et al. Are the estimated needs for mental health care among adolescents from different socioeconomic backgrounds met equally in Sweden? A longitudinal survey-registry linkage study. Eur Child Adolesc Psychiatry 2024 Aug;33(8):2581-2591. [doi: 10.1007/s00787-023-02341-2] [Medline: 38157010]
9. Ojansivu I, Kettunen K, Alajoutsijärvi K. At the temporary-permanent interface: overcoming knowledge boundaries with boundary objects. Scand J Manag 2021 Jun;37(2):101150. [doi: 10.1016/j.scaman.2021.101150]
10. Star SL, Griesemer JR. Institutional ecology, `Translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Soc Stud Sci 1989 Aug;19(3):387-420. [doi: 10.1177/030631289019003001]
11. Leigh Star S. This is not a boundary object: reflections on the origin of a concept. Sci Technol Human Values 2010 Sep;35(5):601-617. [doi: 10.1177/0162243910377624]
12. Fominykh M, Prasolova-Førland E, Divitini M, Petersen SA. Boundary objects in collaborative work and learning. Inf Syst Front 2016 Feb;18(1):85-102. [doi: 10.1007/s10796-015-9579-9]
13. Black LJ. When visuals are boundary objects in system dynamics work. Syst Dyn Rev 2013 Apr;29(2):70-86. [doi: 10.1002/sdr.1496]
14. Heer J, Agrawala M. Design considerations for collaborative visual analytics. Inf Vis 2008 Mar;7(1):49-62. [doi: 10.1057/palgrave.ivs.9500167]
15. Reich Y, Konda SL, Monarch IA, Levy SN, Subrahmanian E. Varieties and issues of participation and design. Design Studies 1996 Apr;17(2):165-180. [doi: 10.1016/0142-694X(95)00000-H]
16. Neo4j graph database. neo4j. 2024. URL: https://neo4j.com/product/neo4j-graph-database/ [accessed 2024-04-16]

XSL•FO
RenderX

17. SCB. SCB Statistikdatabasen. 2024. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/ [accessed 2024-04-16]
18. NeoDash - dashboard builder for neo4j. neo4j. URL: https://neo4j.com/labs/neodash/ [accessed 2025-06-16]
19. Amberscript. Amberscript Global BV. URL: https://www.amberscript.com/en/ [accessed 2025-06-16]
20. Yang Z, Zhang H, Yuan Z, Oja E. Kullback-leibler divergence for nonnegative matrix factorization. In: Honkela T, Duch W, Girolami M, Kaski S, editors. Artificial Neural Networks and Machine Learning – ICANN 2011: Springer; 2011:250-257. [doi: 10.1007/978-3-642-21735-7_31]
21. Lefevre A, Bach F, Fevotte C. Online algorithms for nonnegative matrix factorization with the itakura-saito divergence. Presented at: 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA); Oct 16-19, 2011 p. 313-316. [doi: 10.1109/ASPAA.2011.6082314]
22. scikit learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html [accessed 2024-05-08]
23. CountVectorizer. scikit Learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html#sklearn.feature_extraction.text.CountVectorizer.transform [accessed 2024-05-08]
24. Scikit Learn Decomposition NMF. URL: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html#sklearn.decomposition.NMF [accessed 2024-05-08]
25. Scikit Learn MiniBatch NMF. URL: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.MiniBatchNMF.html [accessed 2024-05-08]
26. Jensen JK, Alkan Olsson J, von Post M, Isaksson C. Reconnecting children to nature: the efficacy of a wildlife intervention depends on local nature and socio‐economic context, but not on urbanisation. People Nature 2024 Oct;6(5):1987-2000 [FREE Full text] [doi: 10.1002/pan3.10702]

Abbreviations

NLTK: Natural Language Toolkit
OECD: Organisation of Economic Co-operation and Development
WHO: World Health Organization

Edited by E Mensah; submitted 11.09.24; peer-reviewed by P Cooke, T Sterner; revised version received 29.04.25; accepted 08.05.25; published 28.07.25.

Please cite as:
Krishna H, Darwich AS, Meijer S
Issues in Identifying Strategies for Youth Mental Well-Being in Stockholm Municipalities Using Participatory Sessions and Text Mining: Qualitative Study
Online J Public Health Inform 2025;17:e66377
URL: https://ojphi.jmir.org/2025/1/e66377
doi:10.2196/66377

© Harsha Krishna, Adam S Darwich, Sebastiaan Meijer. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 28.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

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

Kaori Yamaguchi[1], PhD; Takemi Akahane[2], MD, PhD; Emi Yasuda[1], PhD; Manabu Akahane[1], MD, PhD

[1]Department of Health and Welfare Services, National Institute of Public Health, 2-3-6, Minami, Wako City, Japan

[2]Department of Gastroenterology, Nara Medical University, Kashihara, Japan

Corresponding Author:
Kaori Yamaguchi, PhD
Department of Health and Welfare Services, National Institute of Public Health, 2-3-6, Minami, Wako City, Japan

Abstract

Background: The outbreak of COVID-19 in 2019 led governments worldwide to introduce various public health measures, which included restrictions on travel and public gatherings, effectively reducing the spread of the virus and associated mortality rates. In Japan, nonlegally binding restrictions on outings effectively curbed infections, as in other countries. However, the restrictions impacted lifestyles, including reduced physical activity, increased frailty, and overeating issues, beyond the effect of preventing the spread of infection. Various factors such as personality, age, and cultural norms influenced outing behavior during the pandemic, which varied by activity type.

Objective: To elucidate the association between personality traits and changes in outing behaviors during the COVID-19 pandemic, as well as to clarify age-specific differences in outing behaviors, focusing on different types of outings.

Methods: A cross-sectional survey was conducted using a web-based questionnaire in January 2021, when Japan announced its second emergency declaration during the pandemic. Overall, 1236 participants were recruited, with an equal number of participants for each gender and 10-year age group. The survey included questions regarding changes in the frequency of three types of outings—medical institution visits, eating out, and traveling—in addition to participants' personality traits, such as sociability and morality. Multinomial logistic regression analysis was performed to analyze the association between personality traits and changes in different outing behaviors. Stratified analysis by age group was also performed.

Results: The findings revealed that 790 participants reported no change in medical institution visits, although the frequency of eating out and traveling decreased during the pandemic. Regarding an age-wise comparison, a higher percentage of older people reported no change in medical institution visits but reported a decrease in eating out and traveling than younger people. Multinomial logistic regression analysis stratified by age showed that sociable people were more likely to report a decrease in the frequency of medical institution visits and an increase in the frequency of eating out (odds ratio [OR] 1.92, 95% CI 1.36 - 2.71, P<.001; OR 2.57, 95% CI 1.19 - 5.54, P=.016, respectively), and participants with a strong sense of responsibility were more likely to report a decrease in the frequency of traveling (OR 1.76, 95% CI 1.14 - 2.72, P=.011) among younger adults. Among older adults, strongly responsible individuals were less likely to eating out frequently (OR 2.56, 95% CI 1.12 - 5.82, P=.026).

Conclusions: We examined various behavioral changes observed during the pandemic for different types of outings and their associations with personality traits, as well as differences between age groups. The findings could help promote an understanding of how to effectively communicate and engage in appropriate behaviors in public health emergency settings.

(Online J Public Health Inform 2025;17:e63120) doi:10.2196/63120

KEYWORDS

lockdown; personality; outing behaviors; internet survey; public health; Japan

Introduction

Various social measures were implemented to curb the global spread of COVID-19, including maintaining physical distance, hand hygiene, wearing masks, and restrictions on domestic and international travel, with evidence of their effectiveness has been documented [1-3]. For example, Lau et al [4] reported that the doubling time of infections increased from 2 to 4 days after a lockdown was imposed in Wuhan, China. Fowler et al [5] reported that in the United States, a stay-at-home order was associated with a more than 30% reduction in weekly COVID-19 cases after 1 week and an approximately 60% reduction in weekly fatalities after 3 weeks in localities that implemented the measure.

XSL•FO
RenderX

Although travel and public gathering restrictions effectively controlled the spread of infection, they led to some unintended effects on the lifestyles of adults, especially older adults, resulting in health impairments [6-10]. For instance, Shinohara conducted a prospective six-month cohort study immediately after the first travel and public gathering restrictions in Japan, reporting an increase in frailty among older people [6]. Esquinas reported a decrease in physical activity and an increase in sedentary lifestyle for community-dwelling older people several weeks after a lockdown was implemented in Spain [7]. In a survey conducted after the implementation of restrictions accompanying the pandemic in Slovakia, Lorková et al [8] reported changes in eating habits, demonstrating that more than 30% of Slovak adults, both men and women, indulged in overeating, and more than half reported weight gain. Tavolacci et al [9] reported that the proportion of students engaging in physical activities decreased, and the risk of depression increased for men among students of a French university. Moreover, unemployment and reduced income during the lockdown affected the economic and psychological well-being of young adults in 6 countries across different regions [10].

Restrictions on outings also impacted disease prevention efforts, with the number of people receiving medical screenings decreasing during the pandemic [11-13]. Bakouny et al [11] conducted a hospital survey in Massachusetts during the pandemic and found that the number of patients receiving cancer screening and diagnosis had decreased in the previous 3 months compared with the 3 months before the pandemic. A trend analysis in Spain also revealed a decline in the number of cancer diagnoses during and after the lockdown among youth, adults, and older adults at primary care clinics [12]. Furthermore, a study in Japan found a decrease in the number of patients receiving cancer screening under the restrictions and its association with anxiety triggered by COVID-19 [13]. Such trends are concerning because delays in disease detection and treatment could lead to increased mortality rates in the long term [14].

Restrictions on outings in Japan were aimed at preventing the spread of infection by reducing person-to-person contact and, consequently, large-scale cluster infections at specific locations or events with large groups of people, which would have further disrupted the medical system [15,16]. The restrictions were not legally binding but rather represented a declaration of emergency urging citizens to change their outing behaviors to prevent the spread of infections. This was implemented multiple times (each time lasting from 1 to 3 weeks) between 2020 and 2021 in all prefectures across the country, which effectively helped curtail infections [17]. Although the declaration was not legally binding, Japanese people strictly adhered to the restrictions on going out, and the streets were largely deserted.

Given the wide impact of behavioral restrictions beyond preventing the spread of infection, their appropriate implementation is important. For implementation to be adequate, it must be recognized that people's preventive behavior is influenced by factors such as their health and information literacy, gender, age, and personality traits [18-30].

Personality traits have been reported to be associated with various types of behaviors related to the pandemic [25-28,30]. Previous studies have highlighted the association between personality traits, such as morality, conscientiousness, and neuroticism, and COVID-19 prevention behavior, such as maintaining hygiene and social distancing [25-27]. There have also been reports on the association of outing behaviors with personality traits such as modesty, honesty, sociability, and extraversion [23,28,29]. Additionally, studies have highlighted the relationship between preventive behavior and age [20-22,30]. Older adults positively adopted preventive measures and had higher health-related literacy [20,21]. Furthermore, among older people, unlike younger people, perceptions of the seriousness of the situation were associated with preventive behavior [22].

Preventive behavior is also influenced by culture, as Abuliezi et al [31] reported, Japanese medical students performed significantly better during prevention measures than their American counterparts. Previous studies have reported that the Japanese are sensitive to social pressure, have a strong tendency to avoid negative evaluations from others, and adhere to infection-prevention protocols, driven by social norms and cultural background [31-34].

Furthermore, regarding outing behaviors during the pandemic, different behavioral changes were reported depending on the type of outing. For instance, eating out and leisure activities decreased, whereas visits to neighborhood parks increased [23,24,35,36]. These findings indicate that the impact of the pandemic on outings may have varied depending on the type of activity.

Preventive behavior, including restrictions on outings, is influenced by various factors such as personality, age, and culture, with behavior also differing depending on the type of outing; however, no findings considering them collectively have been reported. Therefore, we aimed to clarify the association between personality traits and changes in outing behaviors during the COVID-19 pandemic, as well as determine age-specific differences, focusing on three types of outings: medical institution visits, eating out, and traveling. The findings could serve as useful tools when considering effective implementation that takes into account population characteristics for preparedness for emergencies such as infectious disease epidemics.

Methods

Study Design, Setting, and Participants

A cross-sectional analysis was conducted using a web-based questionnaire survey on January 27 - 29, 2021, when the COVID-19 pandemic was ongoing in Japan. The central government had declared a state of emergency during the survey period in 11 of the 47 prefectures (Tochigi, Saitama, Chiba, Tokyo, Kanagawa, Gifu, Aichi, Kyoto, Osaka, Hyogo, and Fukuoka) according to the number of infections, situation of the medical system, and the level of development of monitoring systems [15-17]. It was the second declaration, 8 months after the first emergency declaration in Japan [15-17].

XSL•FO
RenderX

An internet-based research company, Macromill, Inc, Japan was engaged to recruit participants and collect responses and the company's registered monitors include over 30 million individuals living in Japan. The survey included 1236 participants aged between 20 and 79 years from across Japan. The company was responsible for recruiting participants and collecting responses. The participants were randomly selected, and an email was sent inviting them to take the survey. An equal number of participants from both genders and 10-year age groups were recruited, with registrations closing once the targeted sample size was achieved. The participants completed the questionnaires through email and received a small reward, as determined by the company.

Data Collection

Cross-sectional data from the web-based questionnaire included information on participants' backgrounds, health-related items, changes in outing behaviors due to the pandemic, and personality traits (Table 1). Economic status was reported to influence outing behavior during the pandemic [37], and the analysis included participants' income as an adjustment variable. Changes in the frequency of three types of outings were included: medical institution visits, eating out, and traveling. Scoring was performed on a five-point Likert scale (1=Increased, 2=Somewhat increased, 3=No change, 4=Somewhat decreased, 5=Decreased).

XSL•FO
RenderX

Table . Items included in the questionnaire

Question[a]	Response
What is the number of people, including yourself, living in your house?	1: 1, 2: 2, 3: 3, 4: 4, 5: 5, 6: 6 or more
Please indicate the family members living with you. (Multiple answers allowed).	1: Infants or toddlers (not attending nursery or kindergarten) 2: Infants or toddlers (attending nursery or kindergarten) 3: Students (elementary school–junior high school) 4: Students (senior high school–college, including vocational school) (attending school in the same prefecture) 5: Students (senior high school–college, including vocational school) (attending school in another prefecture) 6: Adults 7: Older adults (youngest-old: aged 65 - 74) 8: Older adults (middle-old: above 75) 9: Other (please specify)
Please tell us about your chronic disease(s). Choose from options 1 - 6 above for all except "Other."	
Hypertension	1: Undergoing treatment (oral medication), 2:Follow-up (regular hospital visits only), 3: Follow-up (without hospital visits), 4: Untreated, 5: Treated, 6: Not applicable.
Diabetes mellitus	Same as above
Heart disease	Same as above
Cerebrovascular disease	Same as above
Malignant tumor (cancer, leukemia, etc)	Same as above
Respiratory disease (asthma, emphysema, etc)	Same as above
Dyslipidemia (hyperlipidemia)	Same as above
Liver disease	Same as above
Dementia	Same as above
Mental illness (depression, etc)	Same as above
Other (please specify)	Same as above
Please tell us about your current condition.	1: I am physically healthy, 2: I am mentally healthy, 3: I am financially stable.
Please tell us about yourself.	
I am diligent.	1: Strongly disagree, 2: Disagree, 3: Somewhat disagree 4: Somewhat agree, 5: Agree, 6: Strongly agree
I am sociable.	Same as above
I have a strong sense of responsibility.	Same as above
I have a strong sense of morality.	Same as above
I am cooperative.	Same as above
I am honest.	Same as above
Please tell us about the changes you experienced in the following compared with last year.	
Medical institution visit frequency	1: Increased, 2: Somewhat increased, 3: No change, 4: Somewhat decreased, 5: Decreased
Eating out frequency	Same as above
Traveling frequency	Same as above

[a] age, gender, marital status, employment status, yearly household income, and prefectures were prerecorded during the members' registration with the survey panel.

XSL•FO
RenderX

The survey referred to the Japanese version of the Personality Inventory, which is based on the "Big Five" model to assess personality traits [38]. Japanese personality traits "sense of morality," "diligence," and "strong sense of responsibility" were also included because these could influence preventive behaviors toward COVID-19 [26,39]. To minimize the burden on participants, the items were grouped into six categories: "diligent," "sociable," "cooperative," "honest," "strong sense of responsibility," and "strong sense of morality." Regarding the response, the original seven-point Likert scale was modified to a six-point scale for aggregation and analysis purposes [15,40]. The participants were asked to indicate their agreement levels for each personality trait using the scale (1=Strongly disagree, 2=Disagree, 3=Somewhat disagree, 4=Somewhat agree, 5=Agree, 6=Strongly agree).

Data Analysis

A multinomial logistic regression was performed to analyze the association between personality traits and changes in outing behaviors. After verifying multicollinearity, all independent variables representing personality traits were simultaneously inputted. Additionally, a stratified analysis was performed by age group. We classified participants into two groups: younger and older adults. Considering the number of participants in each age group, the older group included individuals aged 60 and above.

Regarding the variables, the scoring for each outing was classified into three groups—increase (1=Increased and 2=Somewhat increased), no change (3=No change), and decrease (4=Somewhat decreased and 5=Decreased)—which were used as dependent variables, and "no change" was used as the baseline category for multinomial logistic regression. The responses for personality traits were categorized into two groups—applicable (4=Somewhat agree, 5=Agree, and 6=Strongly agree) or not applicable (1=Strongly disagree, 2=Disagree, and 3=Somewhat disagree)—which were considered as independent variables. The following variables were included as covariates: gender, age, marital status, living alone, living with children (aged younger than 15 years), living with older adults (aged older than 65 years), low household income (less than JPY 4 million (USD 36500) annually), self-reported disease, self-reported health, employment status, and emergency declaration status.

For all statistical analyses, we used STATA 17/SE (StataCorp LLC) with a statistical significance of $P<.05$.

Ethical Considerations

This study was approved by the Ethics Committee of the National Institute of Public Health, Japan (NIPH-IBRA#12302, approval date: November 17, 2020), in accordance with the Declaration of Helsinki. All participants provided informed consent for data collection and storage. Written informed consent for study participation was obtained upon registration. The web-based questionnaire survey was conducted by an authorized survey company in adherence with personal information protection regulations. Anonymized data were obtained from the company after survey completion.

Results

Table 2 presents the participants' background characteristics and changes in outing behavior before and during COVID-19. The mean (SD) age of the 1236 participants was 49.4 (16.5) years; 618 (46.6%) were men. In terms of personality traits, over 75% of the participants exhibited attributes such as being "honest" (n=1028, 83.2%), having a "strong sense of responsibility" (n=957, 77.4%), and having a "strong sense of morality" (n=965, 78.1%).

XSL•FO
RenderX

Table . Characteristics of participants and changes in outing behaviors.

Characteristics	All participants n=1326	Younger adults n=824	Older adults (older than 60 years) n=412
Sociodemographic factors			
Age (years), mean (SD)	49.4 (16.5)	39.9 (11.1)	68.3 (5.2)
Gender (men), n (%)	618 (46.6)	412 (50.0)	206 (50.0)
Annual household income (<JPY 4 million (USD 36500)), n (%)	364 (29.5)	198 (24.0)	166 (40.3)
(unknown), n (%)	275 (22.3)	217 (26.3)	58 (14.1)
Employment (current), n (%)	749 (60.6)	602 (73.1)	147 (35.7)
Marital status (married), n (%)	446 (36.1)	367 (44.5)	79 (19.2)
Living alone, n (%)	207 (16.8)	149 (18.1)	58 (14.1)
Living with children aged 15 years and younger, n (%)	267 (21.6)	256 (31.1)	11 (2.7)
Living with older adults aged 65 years and older, n (%)	392 (31.7)	148 (18.0)	244 (59.2)
Self-reported disease (more than one disease in treatment), n (%)	542 (43.9)	268 (32.5)	274 (66.5)
Self-reported health (good), n (%)	928 (75.1)	606 (73.5)	322 (78.2)
Emergency declaration[a] when conducting the survey (in progress), n (%)	783 (63.4)	506 (61.4)	277 (67.2)
Personality traits			
Diligent (applicable), n(%)	709 (57.4)	412 (50.0)	297 (72.1)
Sociable (applicable), n(%)	554 (44.8)	357 (43.3)	197 (47.8)
Having a strong sense of responsibility (applicable), n(%)	957 (77.4)	594 (72.1)	363 (88.1)
Having a strong sense of morality (applicable), n(%)	965 (78.1)	596 (72.3)	369 (89.6)
Cooperative (applicable), n(%)	887 (71.8)	547 (66.4)	340 (82.5)
Honest (applicable), n (%)	1028 (83.2)	644 (78.2)	384 (93.2)
Frequency of outing behaviors			
Medical institution visits, n (%)			
Increased	88 (7.1)	68 (8.3)	20 (4.9)
Remained unchanged	790 (63.9)	496 (60.2)	294 (71.4)
Decreased	358 (29.0)	260 (31.6)	98 (23.8)
Eating out, n (%)			
Increased	52 (4.2)	45 (5.5)	7 (1.7)
Remained unchanged	311 (25.2)	217 (26.3)	94 (22.8)
Decreased	873 (70.6)	562 (68.2)	311 (75.5)
Traveling, n (%)			
Increased	28 (2.3)	25 (3.0)	3 (0.7)
Remained unchanged	273 (22.1)	201 (24.4)	72 (17.5)
Decreased	935 (75.7)	598 (72.6)	337 (81.8)

[a] Following the outbreak of COVID-19, a state of emergency was declared in Japan during the survey period in 11 of the 46 prefectures: Tochigi, Saitama, Chiba, Tokyo, Kanagawa, Gifu, Aichi, Kyoto, Osaka, Hyogo, and Fukuoka.

XSL•FO
RenderX

Regarding changes in outing behavior, participants' responses indicated that the frequency of medical institution visits remained unchanged (n=790, 63.9%). However, the majority of the participants reported a decrease in the frequency of eating out (n=873, 70.6%) and traveling (n=935, 75.6%).

Comparing younger and older age groups, a higher proportion of older individuals reported no change in medical institution visits, while a higher proportion of younger individuals reported increased or decreased medical institution visits compared with the older adult group. Regarding eating out and traveling, a lower proportion of younger individuals reported a decrease in eating out and traveling, while a higher percentage reported an increase in eating out and traveling, compared with the older group.

Tables 3 and 4 present the odds ratios (OR) and 95% Confidence Interval (CI) determined through multinomial logistic regression analysis and the stratified analysis for younger and older adults. The analysis examined the association between six personality traits and changes in the frequency of outing behaviors, such as increase, no change, and decrease, in three different types of outings and those for younger and older age groups. Participants with each of these traits were compared with those without these personality traits.

XSL•FO
RenderX

Table . Association between personality traits and changes in outing behaviors (among all participants, n=1326).

Variable	Odds ratio (95% CI)[a]	*P* value
Medical institution visits		
Diligent		
Increase	0.71 (0.41-1.23)	.223
Decrease	0.93 (0.69-0.69)	.623
Sociable		
Increase	0.86 (0.51-1.46)	.582
Decrease	1.55 (1.17-2.06)	.002
Responsibility		
Increase	1.56 (0.86-2.83)	.144
Decrease	1.17 (0.77-1.77)	.473
Morality		
Increase	0.98 (0.54-1.78)	.960
Decrease	1.39 (0.93-2.08)	.112
Cooperative		
Increase	1.34 (0.73-2.49)	.346
Decrease	0.91 (0.65-1.26)	.561
Honest		
Increase	0.60 (0.33-1.10)	.098
Decrease	1.08 (0.71-0.71)	.714
Eating out		
Diligent		
Increase	1.13 (0.51-2.52)	.763
Decrease	1.05 (0.77-1.42)	.774
Sociable		
Increase	2.04 (1.02-4.08)	.045
Decrease	1.09 (0.81-1.47)	.570
Responsibility		
Increase	1.25 (0.50-3.11)	.630
Decrease	1.33 (0.90-1.94)	.149
Morality		
Increase	0.71 (0.34-1.49)	.367
Decrease	0.98 (0.66-1.45)	.922
Cooperative		
Increase	1.07 (0.52-2.21)	.857
Decrease	1.09 (0.78-1.53)	.600
Honest		
Increase	0.94 (0.41-2.14)	.881
Decrease	1.33 (0.91-1.97)	.144
Traveling		
Diligent		
Increase	1.83 (0.65-5.15)	.249
Decrease	1.01 (0.73-1.40)	.936

XSL•FO
RenderX

Variable	Odds ratio (95% CI)[a]	*P* value
Sociable		
Increase	1.39 (0.50-3.84)	.523
Decrease	1.31 (0.96-1.80)	.092
Responsibility		
Increase	1.46 (0.54-3.97)	.456
Decrease	1.81 (1.22-2.68)	.003
Morality		
Increase	0.59 (0.22-1.59)	.298
Decrease	1.41 (0.95-2.08)	.086
Cooperative		
Increase	1.02 (0.39-2.65)	.974
Decrease	1.08 (0.77-1.51)	.656
Honest		
Increase	0.62 (0.24-1.59)	.318
Decrease	1.12 (0.74-1.69)	.587

[a]The odds ratio of decrease in outing frequency was estimated by multinomial logistic regression adjusted for age, gender, household income, employment, marital status, living alone, living with children, living with older adults, self-reported diseases in treatment, self-reported health, and emergency declaration status for all participants and each stratified group at “no change” as baseline outcome. Participants without personality traits were set as the reference for each odds ratio. CI: Confidence Interval

XSL•FO
RenderX

Table . Association between personality traits and changes in outing behaviors (stratified by age).

Variable	Younger adults (n=824) Odds ratio (95% CI)[a]	*P* value	Older adults (>60 years; n=412) Odds ratio (95% CI)[a]	*P* value
Medical institution visits				
Diligent				
Increase	0.76 (0.40-1.44)	.395	0.53 (0.20-1.40)	.198
Decrease	0.98 (0.69-1.39)	.903	0.90 (0.50-1.61)	.714
Sociable				
Increase	1.08 (0.58-2.00)	.818	0.64 (0.23-1.83)	.409
Decrease	1.92 (1.36-2.71)	<.001	1.19 (0.72-1.96)	.509
Responsibility				
Increase	1.29 (0.68-2.47)	.436	4.82 (0.47-49.84)	.187
Decrease	1.07 (0.67-1.72)	.772	1.58 (0.58-4.32)	.371
Morality				
Increase	0.87 (0.44-1.72)	.691	0.98 (0.22-4.30)	.974
Decrease	1.19 (0.76-1.87)	.445	2.38 (0.73-7.78)	.152
Cooperative				
Increase	1.27 (0.63-2.57)	.510	1.61 (0.45-5.77)	.461
Decrease	0.91 (0.63-1.33)	.639	0.89 (0.42-1.88)	.760
Honest				
Increase	0.63 (0.33-1.18)	.148	0.52 (0.09-3.06)	.469
Decrease	1.17 (0.75-1.84)	.482	0.63 (0.20-2.01)	.432
Eating out				
Diligent				
Increase	1.53 (0.67-3.49)	.310	0.24 (0.02-2.73)	.250
Decrease	1.14 (0.79-1.64)	.486	0.81 (0.43-1.55)	.532
Sociable				
Increase	2.57 (1.19-5.54)	.016	1.61 (0.20-13.22)	.655
Decrease	1.13 (0.78-1.63)	.529	1.07 (0.63-1.81)	.811
Responsibility				
Increase	1.13 (0.43-2.96)	.801	—[b]	—
Decrease	1.11 (0.72-1.71)	.627	2.56 (0.12-5.82)	.026
Morality				
Increase	0.57 (0.25-1.30)	.183	—	—
Decrease	1.12 (0.72-1.74)	.619	0.60 (0.22-1.58)	.297
Cooperative				
Increase	0.98 (0.44-2.20	.968	2.35 (0.16-34.42)	.534
Decrease	0.98 (0.66-1.45)	.931	1.53 (0.80-2.93)	.195
Honest				
Increase	1.16 (0.48-2.84)	.738	0.25 (0.01-4.50)	.346
Decrease	1.49 (0.97-2.29)	.066	0.56 (0.18-1.75)	.322
Traveling				
Diligent				

XSL•FO
RenderX

Variable	Younger adults (n=824) Odds ratio (95% CI)[a]	*P* value	Older adults (>60 years; n=412) Odds ratio (95% CI)[a]	*P* value
Increase	2.81 (0.94-8.35)	.063	—	—
Decrease	1.01 (0.69-1.48)	.956	—	—
Sociable				
Increase	1.47 (0.49-4.40)	.486	—	—
Decrease	1.38 (0.95-2.02)	.095	—	—
Responsibility				
Increase	1.14 (0.40-3.29)	.803	—	—
Decrease	1.76 (1.14-2.72)	.011	—	—
Morality				
Increase	0.43 (0.14-1.32)	.139	—	—
Decrease	1.48 (0.95-2.28)	.081	—	—
Cooperative				
Increase	1.11 (0.38-3.18)	.852	—	—
Decrease	1.03 (0.70-1.52)	.874	—	—
Honest				
Increase	0.84 (0.32-2.22)	.721	—	—
Decrease	1.10 (0.70-1.72)	.672	—	—

[a]The odds ratio of decrease in outing frequency was estimated by multinomial logistic regression adjusted for age, gender, household income, employment, marital status, living alone, living with children, living with older adults, self-reported diseases in treatment, self-reported health, and emergency declaration status for all participants and each stratified group at "no change" as baseline outcome (age was adjusted only for all participants). Participants without personality traits were set as the reference for each odds ratio. Values regarding the association between morality and eating out for older adults were not calculated, as the imbalance of the data with an extremely low sample size in a particular category resulted in inappropriate coefficients. The result of the association between personality traits and traveling for the older adult group was not calculated because some categories had extremely low sample sizes, which led to data imbalance, and the model did not converge appropriately. CI: Confidence Interval

[b]not available

In terms of the association between personality traits and outing behaviors, sociable individuals were more likely to report a decrease in the frequency of medical institution visits compared with individuals who were not sociable, for all participants and younger adults (OR 1.55, 95% CI 1.17 - 2.06, *P*=.002 and OR 1.92, 95% CI 1.36 - 2.71, *P*<.001, respectively). Regarding eating out, sociable individuals were more likely to report an increase among all participants and younger adults (OR 2.04, 95%CI 1.02 - 4.08, *P*=.045 and OR 2.57, 95% CI 1.19 - 5.54, *P*=.016, respectively), whereas participants with strong responsibility were more likely to report a decrease among the older adults group (OR 2.56, 95% CI 1.12 - 5.82, *P*=.026). Regarding traveling, participants with strong responsibility were more likely to report a decrease among all participants and younger adults (OR 1.81, 95% CI 1.22 - 2.68, *P*=.003 OR 1.76, 95% CI 1.14 - 2.72, *P*=.001, respectively).

Values regarding the association between morality and eating out for older adults were not calculated, as the imbalance of the data with an extremely low sample size in a particular category resulted in inappropriate coefficients. The result of the association between personality traits and traveling for the older adult group was not calculated because some categories had extremely low sample sizes, which led to data imbalance, and the model did not converge appropriately.

Discussion

Principal Findings

The analysis revealed two main findings. First, the impact of COVID-19 on outing behavior varied depending on the type of activity. Although most participants reported no change in the frequency of medical institution visits, the majority reported a decrease in the frequency of eating out and traveling. To compare the age groups, a higher percentage of older adults reported no change in medical institution visits and a decrease in eating out and traveling compared with younger adults. Second, the impact of COVID-19 on outing behavior was associated with particular personalities. Participants with a sociable personality were more likely to report a decrease in the frequency of medical institution visits and an increase in the frequency of eating out, while participants with a strong sense of responsibility were more likely to report a decrease in the frequency of traveling. The association between changes in outing behavior and personality differed by age group. Younger adults showed similar trends across all participants; meanwhile, older adults with a strong sense of responsibility were more likely to report a decrease in the frequency of eating out.

XSL•FO
RenderX

For visits to medical institutions, previous studies have revealed that seeking emergency treatments remained unchanged and screening examinations decreased during the pandemic [11,12,36,41]. Examining visits to medical institutions, which may include outpatient treatment, we found that this aspect remained unchanged for most participants during the pandemic. Although telemedicine was introduced in Japan in 2018, its application remained limited even during the pandemic; thus, its impact on medical visits was considered minimal, if any. Regarding eating out, studies in Japan [24], Poland [38], and the United States [23] found a reduction in frequency during lockdowns, consistent with our findings. Similarly, for travel, previous studies reported that the COVID-19 pandemic and infection control measures led to a decrease in long-distance travel [42-45], consistent with our findings. Regarding the variation in results by age group, Harrera et al surveyed during the pandemic in Spain and clarified the association between older age and higher frequency of preventive practices, higher health literacy, and greater concern about COVID-19 [22]. They also reported that increased adherence to preventive practices was associated with being female, having greater concern about COVID-19, and seeking more information among the younger generation; meanwhile, for the older generation, it was associated with greater concern about COVID-19 [22]. Concerns about COVID-19 might also have influenced our result that a higher proportion of older adults reduced the frequency of eating out and traveling.

Importantly, three types of outing behaviors were simultaneously investigated, revealing that medical institution visits remained unchanged, whereas eating out and traveling decreased, with the trend differing among age groups. This suggests that the impact of the pandemic on outing behavior varied depending on necessity.

The findings also indicated that outing behaviors in response to COVID-19 were associated with particular personality traits. Personality traits have been reported to be associated with various behaviors related to the COVID-19 pandemic [25,26,29,30]. For example, Qian and Yahara [26] determined an association between preventive behaviors and personality traits such as extraversion, neuroticism, openness, conscientiousness, and agreeableness, as well as a sense of morality. Hygiene and social distancing behaviors have also been associated with personality traits such as extroversion, honesty, modesty, conscientiousness, and neuroticism [25,30]. In a study conducted in Italy, Costantini et al [30] identified personality traits that predicted compliance with COVID-19 prevention measures, including outing behavior. Their findings revealed that the frequency of outing behavior during the pandemic was negatively associated with modesty and creativity and positively associated with liveliness and sociability. The findings on eating out among younger adults align with those of this study. Although the previous study found a positive association between the frequency of outings and sociable personalities, we revealed that sociable people were more likely to report a decrease in the frequency of one type of outing: medical institution visits. This inconsistency may attributed to the evaluations in previous studies focusing on “unnecessary outings” [30].

Cultural aspects were reported to influence behavior during the pandemic. Abuliezi et al [31] reported the characteristics of preventive behavior among Japanese students compared with those in the United States. They determined that Japanese medical students performed significantly better in preventive behavior, although perceived health competence was lower, which means people do not have the confidence to manage their health through health-related habits [31]]. Furthermore, Japanese people tend to avoid social pressure and negative evaluations from others [33,34]. To sum up these cultural aspects, the finding regarding the association between sociable personality and an increase in the frequency of eating out among younger adults may be attributed to peer pressure to eating out or such a trend among peers.

Furthermore, changes in people’s perceptions and behaviors after the pandemic have been reported [46]. Although a better attitude and higher knowledge levels were observed, adherence to protocols remained inconsistent one year after the outbreak of the pandemic among younger adults aged 16 - 30 years in Bangladesh [46]. This survey was also conducted a year after the first emergency declaration; therefore, the results might be influenced by changes in people’s perceptions and attitudes toward the pandemic over time. Considering the results and the survey periods of this study, concerns regarding COVID-19 among older adults might have changed their outing behavior even during the early stages of the pandemic. For a deeper understanding, further analysis would be required to clarify how the experience of the pandemic and repeated emergency declarations affected attitudes and behaviors.

Limitations

Although this study effectively elucidates the association between personality traits and different types of outings during COVID-19 and differences by age groups, there are certain limitations that are worth mentioning. First, the participants responded to the questions based on their memories of the situation prior to the pandemic, and the actual number of outings before and during the pandemic was not observed; thus, response bias is possible. Furthermore, response bias resulting from the high social desirability of Japanese people could have affected the results [47]. Second, given the equal recruitment across age groups and genders, the participants did not represent the general population structure of Japan. In accordance, it should be noted that descriptive statistics regarding changes in the frequency of outing behavior do not represent the situation in Japan. Finally, we cannot eliminate the possibility of selection bias, as we only included participants registered with an internet panel company who had access to the internet and could respond on the web. Nevertheless, panel-based surveys have been widely used in recent years [48-50]. In addition, the participants received a small cash reward for completing the survey, which could have affected the randomness of the sample. However, because the participants were randomly selected from a large number of people registered with the engaged company across Japan, bias is expected to be minimized.

Conclusions

We examined various behavioral changes observed during the pandemic for different types of outings and their associations

XSL•FO
RenderX

with personality traits and differences between age groups. The findings can help in understanding how to effectively communicate and promote appropriate behaviors in public health emergency settings.

Acknowledgments

We wish to thank the internet survey company, Macromill, Inc, Japan. This study was partially supported by a grant-in-aid for Scientific Research from the Ministry of Health, Labour and Welfare, Japan.

Data Availability

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

Authors' Contributions

KY, TA, and MA conceptualized the study. KY conducted the formal analysis and drafted the manuscript. MA acquired funding; MA and TA led the investigation; KY and EY developed the methodology. KY, TA, EY, and MA reviewed and edited the manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. Infection prevention and control in the context of coronavirus disease (COVID-19): a living guideline. 2022. URL: https://iris.who.int/bitstream/handle/10665/353565/WHO-2019-nCoV-ipc-guideline-2022.2-eng.pdf [accessed 2025-05-12]
2. Considerations for implementing and adjusting public health and social measures in the context of COVID-19.: World Health Organization; 2023. URL: https://iris.who.int/bitstream/handle/10665/366669/WHO-2019-nCoV-Adjusting-PH-measures-2023.1-eng.pdf?sequence=1 [accessed 2025-05-12]
3. Girum T, Lentiro K, Geremew M, et al. Optimal strategies for COVID-19 prevention from global evidence achieved through social distancing, stay at home, travel restriction and lockdown: a systematic review. Arch Public Health 2021 Aug 21;79(1):150. [doi: 10.1186/s13690-021-00663-8] [Medline: 34419145]
4. Lau H, Khosrawipour V, Kocbach P, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med 2020 May 18;27(3):taaa037. [doi: 10.1093/jtm/taaa037] [Medline: 32181488]
5. Fowler JH, Hill SJ, Levin R, Obradovich N. Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States. PLOS ONE 2021;16(6):e0248849. [doi: 10.1371/journal.pone.0248849] [Medline: 34111123]
6. Shinohara T, Saida K, Tanaka S, Murayama A, Higuchi D. Did the number of older adults with frailty increase during the COVID-19 pandemic? A prospective cohort study in Japan. Eur Geriatr Med 2021 Oct;12(5):1085-1089. [doi: 10.1007/s41999-021-00523-2] [Medline: 34081313]
7. García-Esquinas E, Ortolá R, Gine-Vázquez I, et al. Changes in health behaviors, mental and physical health among older adults under severe lockdown restrictions during the COVID-19 pandemic in Spain. Int J Environ Res Public Health 2021 Jul 1;18(13):7067. [doi: 10.3390/ijerph18137067] [Medline: 34281004]
8. Lorková M, Gažarová M. Impact of the COVID-19 pandemic on the changes in dietary habits, lifestyle and physical activity in the Slovak population. Rocz Panstw Zakl Hig 2023;74(3):295-307. [doi: 10.32394/rpzh.2023.0268] [Medline: 37577853]
9. Tavolacci MP, Wouters E, Van de Velde S, et al. The impact of COVID-19 lockdown on health behaviors among students of a French university. Int J Environ Res Public Health 2021 Apr 20;18(8):4346. [doi: 10.3390/ijerph18084346] [Medline: 33923943]
10. Li L, Serido J, Vosylis R, et al. Employment disruption and wellbeing among young adults: a cross-national study of perceived impact of the COVID-19 lockdown. J Happiness Stud 2023;24(3):991-1012. [doi: 10.1007/s10902-023-00629-3] [Medline: 36776722]
11. Bakouny Z, Paciotti M, Schmidt AL, Lipsitz SR, Choueiri TK, Trinh QD. Cancer screening tests and cancer diagnoses during the COVID-19 pandemic. JAMA Oncol 2021 Mar 1;7(3):458-460. [doi: 10.1001/jamaoncol.2020.7600] [Medline: 33443549]
12. Coma E, Guiriguet C, Mora N, et al. Impact of the COVID-19 pandemic and related control measures on cancer diagnosis in Catalonia: a time-series analysis of primary care electronic health records covering about five million people. BMJ Open 2021 May 18;11(5):e047567. [doi: 10.1136/bmjopen-2020-047567] [Medline: 34006554]
13. Akahane T, Nakanishi Y, Yoshiji H, Akahane M. Esophagogastroduodenoscopy screening intentions during the COVID-19 pandemic in Japan: web-based survey. JMIR Cancer 2022 Nov 11;8(4):e40600. [doi: 10.2196/40600] [Medline: 36343187]

XSL•FO
RenderX

14. Maringe C, Spicer J, Morris M, et al. The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study. Lancet Oncol 2020 Aug;21(8):1023-1034. [doi: 10.1016/S1470-2045(20)30388-0] [Medline: 32702310]
15. Ministry of Health, Labour and Welfare. About COVID-19. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00001.html [accessed 2024-12-23]
16. Cabinet Agency for Infectious Disease Crisis Management. About COVID-19. URL: https://www.caicm.go.jp/citizen/corona/index.html [accessed 2024-12-23]
17. Cabinet Agency for Infectious Disease Crisis Management. Response based on principle measures policy. URL: https://www.caicm.go.jp/citizen/corona/emergency.html [accessed 2024-12-23]
18. Lermer E, Hudecek MFC, Gaube S, Raue M, Batz F. Early and later perceptions and reactions to the COVID-19 pandemic in Germany: on predictors of behavioral responses and guideline adherence during the restrictions. Front Psychol 2021;12:769206. [doi: 10.3389/fpsyg.2021.769206] [Medline: 34899517]
19. Uddin S, Imam T, Khushi M, Khan A, Ali M. How did socio-demographic status and personal attributes influence compliance to COVID-19 preventive behaviours during the early outbreak in Japan? Lessons for pandemic management. Pers Individ Dif 2021 Jun;175:110692. [doi: 10.1016/j.paid.2021.110692] [Medline: 33526954]
20. Muto K, Yamamoto I, Nagasu M, Tanaka M, Wada K. Japanese citizens' behavioral changes and preparedness against COVID-19: an online survey during the early phase of the pandemic. PLOS ONE 2020;15(6):e0234292. [doi: 10.1371/journal.pone.0234292] [Medline: 32525881]
21. González-Herrera A, Rodríguez-Blázquez C, Romay-Barja M, Falcon-Romero M, Ayala A, Forjaz MJ. Age differences in knowledge, attitudes and preventive practices during the COVID-19 pandemic in Spain. Sci Rep 2022 Dec 2;12(1):20863. [doi: 10.1038/s41598-022-25353-5] [Medline: 36460702]
22. Kim JK, Crimmins EM. Age differences in the relationship between threatening and coping mechanisms and preventive behaviors in the time of COVID-19 in the United States: protection motivation theory. Res Psychother 2020 Dec 31;23(3):485. [doi: 10.4081/ripppo.2020.485] [Medline: 33585294]
23. Cohen JFW, Posluszny H, Falbe J, et al. Restaurant dining during the COVID-19 pandemic among adults with low-income in the United States. Appetite 2022 Jun 1;173:105976. [doi: 10.1016/j.appet.2022.105976] [Medline: 35245643]
24. Parady G, Taniguchi A, Takami K. Corrigendum to travel behavior changes during the COVID-19 pandemic in Japan: analyzing the effects of risk perception and social influence on going-out self-restriction [Transp. Res. Interdiscip. Perspect. 7 (2020) 100181]. Transp Res Interdiscip Perspect 2022 Sep;15:100649. [doi: 10.1016/j.trip.2022.100649] [Medline: 35782787]
25. Abdelrahman M. Personality traits, risk perception, and protective behaviors of Arab residents of Qatar during the COVID-19 pandemic. Int J Ment Health Addict 2022;20(1):237-248. [doi: 10.1007/s11469-020-00352-7] [Medline: 32837433]
26. Qian K, Yahara T. Mentality and behavior in COVID-19 emergency status in Japan: influence of personality, morality and ideology. PLOS ONE 2020;15(7):e0235883. [doi: 10.1371/journal.pone.0235883] [Medline: 32649687]
27. Wu JS, Font X, McCamley C. COVID-19 social distancing compliance mechanisms: UK evidence. Environ Res 2022 Apr 1;205:112528. [doi: 10.1016/j.envres.2021.112528] [Medline: 34953882]
28. Alivernini F, Manganelli S, Girelli L, et al. Physical distancing behavior: the role of emotions, personality, motivations, and moral decision-making. J Pediatr Psychol 2021 Jan 20;46(1):15-26. [doi: 10.1093/jpepsy/jsaa122] [Medline: 33355343]
29. Brouard S, Vasilopoulos P, Becher M. Sociodemographic and psychological correlates of compliance with the COVID-19 public health measures in France. Can J Pol Sci 2020 Dec;53(2):253-258. [doi: 10.1017/S0008423920000335]
30. Costantini G, Di Sarno M, Preti E, et al. Personality in a naturally occurring strong situation: unique effects of HEXACO traits on COVID-19 mitigation. Acta Psychol (Amst) 2022 Apr;224:103529. [doi: 10.1016/j.actpsy.2022.103529] [Medline: 35151956]
31. Abuliezi R, Kondo A, Niitsu K, Ota E. Healthcare graduate students' perceived control and preventive behavior for COVID-19 in Japan and the United States: a cross-sectional study. Front Public Health 2022;10:965897. [doi: 10.3389/fpubh.2022.965897] [Medline: 36388307]
32. Hsu KW, Chao JC, Hsu CY. Environmental risk perception and preventive behavior during the COVID-19 pandemic in central Taiwan. Int J Environ Res Public Health 2021 Sep 21;18(18):9920. [doi: 10.3390/ijerph18189920] [Medline: 34574842]
33. Hashimoto H, Yamagishi T. Two faces of interdependence: harmony seeking and rejection avoidance. Asian J of Social Psycho 2013 Jun;16(2):142-151 [FREE Full text] [doi: 10.1111/ajsp.12022]
34. Hashimoto H, Yamagishi T. Duality of independence and interdependence: an adaptationist perspective. Asian J Soc Psychol 2016 Oct;19(4):286-297 [FREE Full text] [doi: 10.1111/ajsp.12145]
35. Ueno Y, Kato S, Mase T, Funamoto Y, Hasegawa K. Human flow dataset reveals changes in citizens' outing behaviors including greenspace visits before and during the first wave of the COVID-19 pandemic in Kanazawa, Japan. Int J Environ Res Public Health 2022 Jul 18;19(14):8728. [doi: 10.3390/ijerph19148728] [Medline: 35886580]
36. Borges-Machado F, Barros D, Ribeiro Ó, Carvalho J. The effects of COVID-19 home confinement in dementia care: physical and cognitive decline, severe neuropsychiatric symptoms and increased caregiving burden. Am J Alzheimers Dis Other Demen 2020;35:1533317520976720. [doi: 10.1177/1533317520976720] [Medline: 33295781]

XSL•FO
RenderX

37. Wang J, Kaza N, McDonald NC, Khanal K. Socio-economic disparities in activity-travel behavior adaptation during the COVID-19 pandemic in North Carolina. Transp Policy (Oxf) 2022 Sep;125:70-78. [doi: 10.1016/j.tranpol.2022.05.012] [Medline: 35664727]
38. Błaszczyk-Bębenek E, Jagielski P, Bolesławska I, et al. Nutrition behaviors in Polish adults before and during COVID-19 lockdown. Nutrients 2020 Oct 10;12(10):3084. [doi: 10.3390/nu12103084] [Medline: 33050404]
39. Iwasa H, Yoshida Y. Normative data of middle-aged and older Japanese adults for the the Japanese version of the ten-item personality inventory (TIPI-J). Nihon Koshu Eisei Zasshi 2018;65(7):356-363. [doi: 10.11236/jph.65.7_356] [Medline: 30068821]
40. Lee J, Lee SM, Jung E. How important is behavioral change during the early stages of the COVID-19 pandemic? A mathematical modeling study. Int J Environ Res Public Health 2021 Sep 18;18(18):9855. [doi: 10.3390/ijerph18189855] [Medline: 34574785]
41. Wada S. Construction of the big five scales of personality trait terms and concurrent validity with NPI. The Japanese Journal of Psychology 1996;67(1):61-67. [doi: 10.4992/jjpsy.67.61]
42. Cancino RS, Su Z, Mesa R, Tomlinson GE, Wang J. The impact of COVID-19 on cancer screening: challenges and opportunities. JMIR Cancer 2020 Oct 29;6(2):e21697. [doi: 10.2196/21697] [Medline: 33027039]
43. Chen K, Steiner R. Longitudinal and spatial analysis of Americans' travel distances following COVID-19. Transp Res D Transp Environ 2022 Sep;110:103414. [doi: 10.1016/j.trd.2022.103414] [Medline: 35942325]
44. Irawan MZ, Belgiawan PF, Joewono TB, Bastarianto FF, Rizki M, Ilahi A. Exploring activity-travel behavior changes during the beginning of COVID-19 pandemic in Indonesia. Transportation (Amst) 2022;49(2):529-553. [doi: 10.1007/s11116-021-10185-5] [Medline: 33723465]
45. Morar C, Tiba A, Basarin B, et al. Predictors of changes in travel behavior during the COVID-19 pandemic: the role of tourists' personalities. Int J Environ Res Public Health 2021 Oct 24;18(21):11169. [doi: 10.3390/ijerph182111169] [Medline: 34769688]
46. Rupok TA, Dey S, Agarwala R, et al. Knowledge, attitude, practice, and fear level of Bangladeshi students toward Covid-19 after a year of the pandemic situation: a web-based cross-sectional study. PLOS ONE 2023;18(2):e0282282. [doi: 10.1371/journal.pone.0282282] [Medline: 36848394]
47. Heine SJ, Lehman DR. Social desirability among Canadian and Japanese students. J Soc Psychol 1995 Dec;135(6):777-779. [doi: 10.1080/00224545.1995.9713982]
48. Akahane M, Maeyashiki A, Yoshihara S, et al. Relationship between difficulties in daily activities and falling: loco-check as a self-assessment of fall risk. Interact J Med Res 2016 Jun 20;5(2):e20. [doi: 10.2196/ijmr.5590] [Medline: 27323871]
49. Akahane M, Kanagawa Y, Takahata Y, Nakanishi Y, Akahane T, Imamura T. Consumer awareness of food defense measures at food delivery service providers and food manufacturers: web-based consumer survey study. JMIR Form Res 2023 Aug 24;7:e44150. [doi: 10.2196/44150] [Medline: 37616047]
50. Min YH, Lee JW, Shin YW, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014 May 23;16(5):e135. [doi: 10.2196/jmir.3421] [Medline: 24860070]

Abbreviations

OR: odds ratio
CI: Confidence Interval

Edited by E Mensah; submitted 11.06.24; peer-reviewed by S Thawillarp, Y Kotera; revised version received 22.03.25; accepted 18.04.25; published 12.06.25.

Please cite as:
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;17:e63120
URL: https://ojphi.jmir.org/2025/1/e63120
doi:10.2196/63120

© Kaori Yamaguchi, Takemi Akahane, Emi Yasuda, Manabu Akahane. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 12.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is

XSL•FO
RenderX

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

XSL•FO
RenderX

Nowcasting to Monitor Real-Time Mpox Trends During the 2022 Outbreak in New York City: Evaluation Using Reportable Disease Data Stratified by Race or Ethnicity

Rebecca Rohrer, MPH; Allegra Wilson, MSc; Jennifer Baumgartner, MSPH; Nicole Burton, PhD, MPH; Ray R Ortiz, MS, MSW; Alan Dorsinville, MPH; Lucretia E Jones, DrPH, MPH; Sharon K Greene, PhD, MPH

Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, Long Island City, NY, United States

Corresponding Author:
Rebecca Rohrer, MPH
Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, Long Island City, NY, United States

Abstract

Background: Applying nowcasting methods to partially accrued reportable disease data can help policymakers interpret recent epidemic trends despite data lags and quickly identify and remediate health inequities. During the 2022 mpox outbreak in New York City, we applied Nowcasting by Bayesian Smoothing (NobBS) to estimate recent cases, citywide and stratified by race or ethnicity (Black or African American, Hispanic or Latino, and White). However, in real time, it was unclear if the estimates were accurate.

Objective: We evaluated the accuracy of estimated mpox case counts across a range of NobBS implementation options.

Methods: We evaluated NobBS performance for New York City residents with a confirmed or probable mpox diagnosis or illness onset from July 8 through September 30, 2022, as compared with fully accrued cases. We used the exponentiated average log score (average score) to compare moving window lengths, stratifying or not by race or ethnicity, diagnosis and onset dates, and daily and weekly aggregation.

Results: During the study period, 3305 New York City residents were diagnosed with mpox (median 4, IQR 3-5 days from diagnosis to diagnosis report). Of these, 812 (25%) had missing onset dates, and of these, 230 (28%) had unknown race or ethnicity. The median lag in days from onset to onset report was 10 (IQR 7-14). For daily hindcasts by diagnosis date, the average score was 0.27 for the 14-day moving window used in real time. Average scores improved (increased) with longer moving windows (maximum: 0.47 for 49-day window). Stratifying by race or ethnicity improved performance, with an overall average score of 0.38 for the 14-day moving window (maximum: 0.57 for 49 day-window). Hindcasts for White patients performed best, with average scores of 0.45 for the 14-day window and 0.75 for the 49-day window. For unstratified, daily hindcasts by onset date, the average score ranged from 0.16 for the 42-day window to 0.30 for the 14-day window. Performance was not improved by weekly aggregation. Hindcasts underestimated diagnoses in early August after the epidemic peaked, then overestimated diagnoses in late August as the epidemic waned. Estimates were most accurate during September when cases were low and stable.

Conclusions: Performance was better when hindcasting by diagnosis date than by onset date, consistent with shorter lags and higher completeness for diagnoses. For daily hindcasts by diagnosis date, longer moving windows performed better, but direct comparisons are limited because longer windows could only be assessed after case counts in this outbreak had stabilized. Stratification by race or ethnicity improved performance and identified differences in epidemic trends across patient groups. Contributors to differences in performance across strata might include differences in case volume, epidemic trends, delay distributions, and interview success rates. Health departments need reliable nowcasting and rapid evaluation tools, particularly to promote health equity by ensuring accurate estimates within all strata.

(Online J Public Health Inform 2025;17:e56495) doi:10.2196/56495

KEYWORDS

data quality; epidemiology; forecasting; infectious disease; morbidity and mortality trends; mpox; nowcasting; public health practice; surveillance

XSL•FO
RenderX

Introduction

Timeline and Motivation

In 2022, an mpox outbreak occurred in countries where local transmission previously had not been observed, including the United States [1]. New York City was the first urban center in the United States to experience a rapid increase in cases [2]. The first case among New York City residents was diagnosed on May 19, 2022 [3]. The next day, New York City health care providers were notified to immediately report suspected cases to the Provider Access Line at the New York City Department of Health and Mental Hygiene (New York City Health Department) for potential testing through the Public Health Laboratory [3]. On June 21, 2022, the New York City Health Department Incident Command System was activated for a public health response, and on July 8, the New York State Department of Health notified health care providers of the availability of commercial laboratory testing for mpox [4]. The New York City Health Department declared a local state of emergency on August 1 [5], and the Secretary of Health and Human Services declared a nationwide public health emergency on August 4 [6]. As the outbreak subsided, the New York City Health Department partially deactivated mpox emergency response activities on October 31 and fully deactivated these activities on January 31, 2023, aligning with the expiration of the US public health emergency declaration [7].

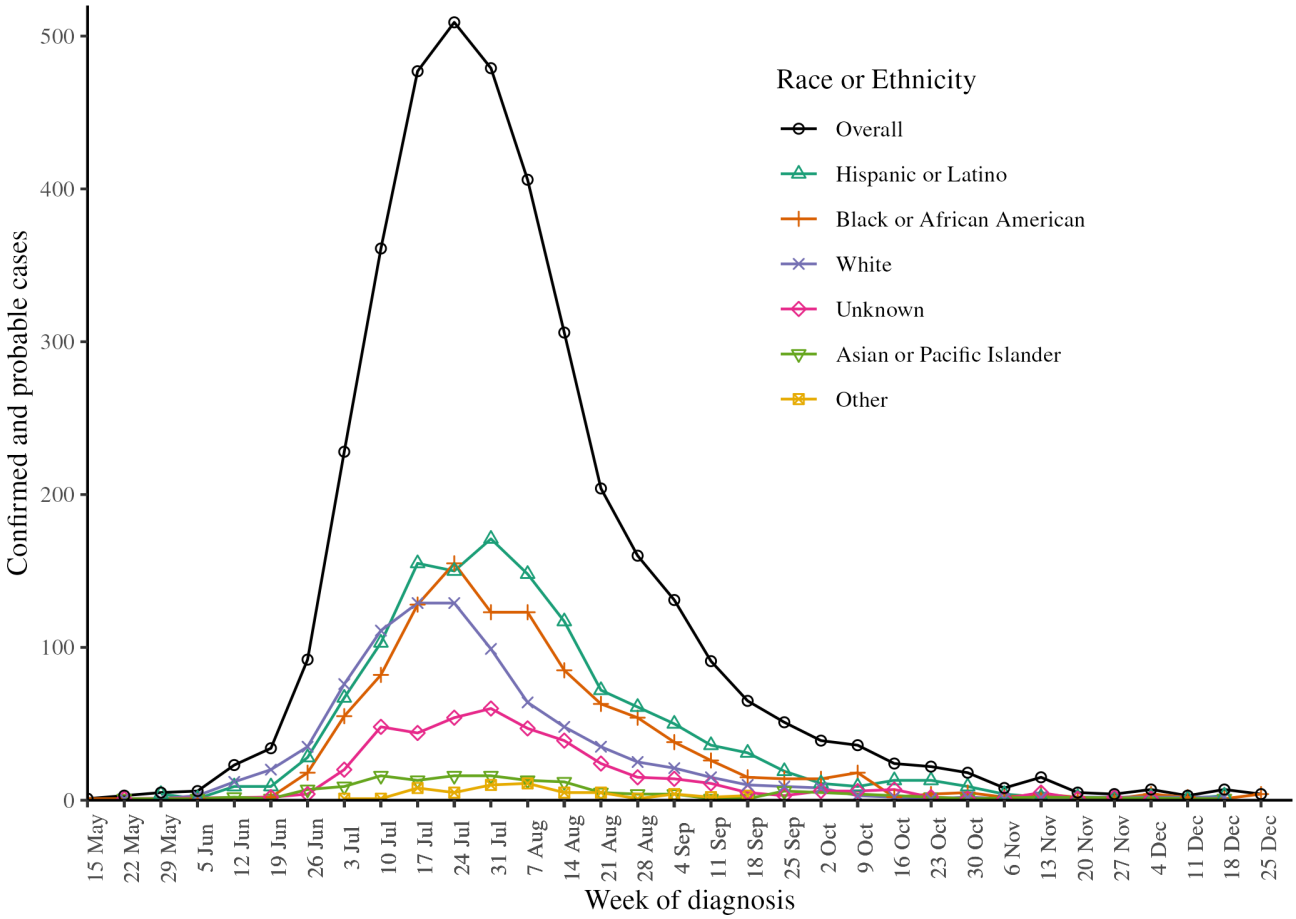
Throughout the emergency response, the New York City Health Department tracked case counts internally and on a public-facing website [8]. Inherent delays (eg, from patient symptom onset to care seeking, laboratory testing, provider and laboratory reporting to the New York City Health Department, and phone interviews with patients to determine the date of onset) make it difficult to interpret recent epidemic trends and make timely decisions during an outbreak. In early August 2022, while reviewing daily epidemic curves with no accounting for data lags, the New York City Health Department leadership inquired whether the outbreak had peaked.

Health Inequities Across Race or Ethnicity Groups

The burden of mpox diagnoses was inequitably distributed by race and ethnicity among patients in the United States [9] and in New York City [10]. Confirmed and probable mpox diagnoses [11] among New York City residents peaked first among White individuals (weeks beginning July 17 and July 24, 2022), then among Black or African American individuals (week beginning July 24, 2022), and then among Hispanic or Latino individuals (week beginning July 31, 2022; Figure 1). Cases then decreased most sharply first among White individuals, then among Black or African American individuals, and then among Hispanic or Latino individuals. Differences in the timing, magnitude, and duration of epidemic peaks by race or ethnicity could reflect, in part, true epidemiologic differences, such as sexual network effects including exposures while traveling early in the outbreak, before local transmission was established, and differences in access to vaccination and treatment [12-16]. In addition, systemic inequities, including heightened stigma, medical mistrust, and inaccessibility of health care services (including financial barriers, inadequate insurance coverage, not having access to a primary care provider, lack of transportation, and lack of convenient care locations) likely contributed to reduced or delayed case ascertainment among Black or African American and Hispanic or Latino individuals [16-21]. Additionally, public health messaging and outreach did not quickly and effectively reach all affected persons, due in part to insufficient accommodation for cultural nuances and linguistic diversity, further contributing to care-seeking delays [17,19,22].

XSL•FO
RenderX

Figure 1. Weekly confirmed and probable mpox cases among New York City residents diagnosed from May through December 2022, overall and stratified by race or ethnicity.



Nowcasting and the COVID-19 Pandemic Precedent

"Nowcasting" refers to predicting the present, and "hindcasting" refers to predicting through the day before the present. Nowcasting and hindcasting methods can be applied to partially accrued reportable disease data to estimate the number of recent events that have not yet been reported [23,24]. Public health agencies have nowcasted various infectious diseases [24-26].

The New York City Health Department first used nowcasting to improve real-time situational awareness during the COVID-19 pandemic public health emergency [24], applying a method called Nowcasting by Bayesian Smoothing (NobBS) [23,27]. NobBS requires a case line list of "date of interest" and "report date" to assess the past delay distribution and epidemic trend and projects the number of cases during a user-specified moving window ending on a date representing "now" [23].

We applied lessons learned from an evaluation of nowcasting COVID-19 [24] to mpox, including (1) using a negative binomial distribution instead of the NobBS default Poisson distribution, (2) using a 2-week moving window length for diagnoses, and (3) removing the display of estimates of diagnoses on weekends, given lack of adjustment for day-of-week effects. Additionally, we wished to nowcast mpox by onset date and to stratify by race or ethnicity, neither of which was previously implemented for COVID-19 at the New York City Health Department. We sent daily automated nowcasting reports to surveillance data leadership starting September 19, 2022; implementation delays were driven by complexities in determining the onset report date and limited staff resources. To monitor differences in epidemic trends across groups, we started stratifying nowcasts by race or ethnicity on September 29.

Objectives

First, we documented challenges in developing input files for daily hindcasts of mpox cases among New York City residents by diagnosis date and by onset date, overall and stratified by race or ethnicity. Our goal was to provide methodologists developing nowcasting tools with greater insight into how relevant data are collected locally during a public health emergency. Second, we conducted a retrospective evaluation of hindcasting performance for New York City residents diagnosed with confirmed or probable mpox [11] from July 8 through September 30, 2022, capturing the outbreak peak and decline, compared with fully accrued case counts as of September 1, 2023. We used a 14-day moving window for hindcasting by diagnosis date and a 21-day moving window for hindcasting by onset date in real time and assessed whether other moving window lengths or a weekly time unit would have performed better. Third, we assessed mpox hindcast accuracy when stratifying by race or ethnicity.

XSL•FO
RenderX

Methods

Data Collection

We used onset, diagnosis, and reporting dates, as well as race and ethnicity data from the New York City Health Department's mpox surveillance database. Reports were imported electronically from laboratories via the New York State Electronic Clinical Laboratory Reporting System [28,29] and from health care providers via Reporting Central, through the electronic Universal Reporting Form [30]. Information from providers reporting by phone was entered into the surveillance database via on-call physician notes. We included patients who tested positive for either mpox virus (confirmed cases) or orthopoxvirus (probable cases), as detailed in standard case definitions [11].

The Surveillance and Investigations Unit of the Mpox Emergency Response Team at the New York City Health Department conducted patient phone interviews as soon as possible after the initial report of diagnosis to determine risk factors for exposure, identify contacts, and prevent further transmission. These interviews included questions on symptom onset date, self-reported race and ethnicity, and recent history of sexual contact. Responses were entered into the surveillance database.

Data Point Selection

We selected the relevant "dates of interest" (diagnosis or onset date) and their respective report dates. The diagnosis date was defined as the specimen collection date of the first positive laboratory test, which was ascertained via electronic laboratory reporting. The symptom onset date for mpox illness was elicited during the patient interview and manually entered. The respective report dates were the different dates that the New York City Health Department ascertained the dates of interest. The diagnosis report date was defined as the date the first positive laboratory result indicating a patient met confirmed or probable case criteria [11] was received by the New York City Health Department. The onset report date was calculated based on the source for establishing the onset date, which was most commonly patient interview (Table 1).

We reviewed cases with long (≥50 days) or negative spans between date of interest and its report date to identify cases requiring additional data cleaning. Patients with a missing onset date were excluded from onset nowcasting. Of 2493 patients diagnosed during the study period and with an available onset date, 2099 (84%) had different report dates for diagnosis and onset, with a median of 2 (IQR 1 - 4) days between diagnosis report date and onset report date.

Table . Mpox onset report date sources, in descending order of preference, as available from the New York City Health Department's surveillance database and as used for Nowcasting by Bayesian Smoothing.

Onset date source	Onset report date source	Onset report date source for 2278 patients with an available onset date from July 8 through September 30, 2022, n (%)
Health care provider report, where onset date on form matches mpox onset date in case record	Electronic universal reporting form receipt date	35 (1.5)
Patient interview	Interview date	2038 (89.5)
Administrative log	Date administrative interview log was changed for the final time from "Assigned" to another status, for example, "Complete" or "Sent to supervisor for review"[a]	23 (1.0)
Any source, if onset before August 1 or outlier in quality assurance review	Manually hard-coded based on free-text notes in the surveillance database	49 (2.2)
Any source, if no other date available, or if later than the date set earlier in the hierarchy	Date the case was first set as confirmed or probable	133 (5.8)

[a]Applied to patients with onset starting August 1, 2022. Before then, interview dates were likely to be reported in on-call physician notes only, and assigning the onset report date based on the interview log would have been inaccurate.

On September 1, 2023, we created a frozen analytic line list of mpox cases among New York City residents with the minimum necessary variables to evaluate nowcasting performance—diagnosis date, diagnosis report date, onset date, onset report date, and race or ethnicity. This dataset was separately filtered by patients with diagnosis (n=3305) or known illness onset (n=2278) during the study period, from July 8 through September 30, 2022. We started the study period on July 8, 2022, when commercial laboratory testing became available, and ended on September 30, 2022, because case counts were sparse thereafter (Figure 1).

We characterized the delay distribution from diagnosis to diagnosis report and from onset to onset report by median number of days, IQR, and 90th percentile. We assessed delays overall during the study period and stratified by month and race or ethnicity. We used Kruskal-Wallis tests to assess whether delay distributions varied across race or ethnicity.

Retrospective Nowcasting Evaluation

We mimicked prospective surveillance on Wednesdays for case counts through Tuesdays by using the R package NobBS (The R Foundation) [27] and restricting to data that had been available at the time. We evaluated hindcast performance across moving window length (14, 21, 28, 35, 42, and 49 days and 2, 3, 4, 5, 6, and 7 weeks), time unit (day vs week), and stratification (overall or stratified by race or ethnicity). For the maximum

XSL•FO
RenderX

delay value, we used the NobBS default of the moving window length minus 1.

We chose to mimic surveillance on Wednesdays to balance operational constraints. Hindcast estimates produced on Mondays and Tuesdays could be underestimated because of reduced care-seeking and laboratory reporting on weekends, and hindcasts conducted on Thursdays and Fridays might be received by decision makers too late in the work week to affect that week's planned public health actions.

To evaluate moving window lengths at daily resolution, we retained the number of estimated cases for each of the prior 7 days. For weekly resolution, we aggregated cases to 7-day periods and retained the estimate for the most recent week. While data from diagnoses on all days of the week were included in model inputs, when conducting the performance evaluation, we evaluated only daily diagnosis estimates from weekdays. This was because diagnoses were reduced on weekends when health care provider availability was more limited. This exclusion did not apply to estimates of onsets or weekly time periods.

Each window length was assessed for periods ending on Tuesdays once the number of days or weeks of that window length had elapsed since the July 8, 2022, start date. For example, we assessed the performance of a 14-day moving window beginning the 14-day period from July 13 through 26, 2022, shifting forward 1 week from July 20 through August 2, and continuing to shift forward 1 week at a time until ending with the period from September 14 through 27, 2022, for a total of 10 models run. For each model, we retained diagnoses for the last 7 days in the window, then excluded weekends, for a total of 50 estimates (5 weekdays from each of 10 models with different end dates). These 50 estimates were used for the performance evaluation. Scenarios with longer moving windows or with weekly aggregation had fewer estimates available for evaluation.

When stratifying by race or ethnicity, we used the "strata" option in NobBS. This option estimated the delay distribution across all race and ethnicity groups and the epidemic curves separately for each group. These analyses were restricted to Black (including African American or Afro-Caribbean), Hispanic or Latino, and White patients because of low case counts in other groups, including Asian, Native Hawaiian or Pacific Islander, and Native American or Alaska Native. We suspected the delay distribution could vary across race or ethnicity groups given differential access to diagnosis and accessibility for interviews, motivating us to compare the accuracy of stratified and unstratified estimates.

For each date of interest (ie, diagnosis or onset), we evaluated groups of estimates—moving window lengths against the lengths used in real time, stratified estimates, and weekly versus daily time units. Drawing from prior evaluations, we evaluated hindcasting performance using the log score [23], mean absolute error (MAE) [24,31,32], relative root mean square error (rRMSE) [24], and 95% prediction interval (PI) coverage [24,32].

We used the log score to evaluate the accuracy of the posterior predictive distribution of each hindcast [23]. We assigned predictive distributions to bins of possible values of fully accrued case counts. For unstratified hindcasts, we used bin widths of 10 cases ranging from 0 - 99 for daily hindcasts and of 50 cases ranging from 0 - 549 for weekly hindcasts. For stratified hindcasts, we used bin widths of 5 cases ranging from 0 - 39 for daily hindcasts and of 20 cases ranging from 0 - 179 for weekly hindcasts. These bin widths were selected to yield similar numbers of bins (10, 11, 8, or 9 bins, respectively), to enable comparisons across scenarios with widely varying case volumes. The log score was the natural log of the probability assigned to the bin in which the true count fell [23]. If the probability assigned to the bin for the true count was 0, then we assigned a lower limit log score of −10; this was necessary for only one estimate, for hindcasting for August 23, 2022, by week of onset using a 4-week moving window, stratified among Hispanic or Latino patients. We calculated the average log score across all days or weeks retained for evaluation. We report the exponentiated average log score (ie, average score), which is the average probability NobBS assigned to the bin containing the true case count [23]. Higher average scores indicated more accurate performance.

We also calculated the daily or weekly MAE and average daily or weekly rRMSE across all individual days or weeks evaluated to compare point estimates of hindcasted cases with the final number of cases reported after data accrued. Lower MAE and lower rRMSE indicated better performance, with estimates closer to final counts. MAE is dependent on case volume, making it useful for comparing scenarios with similar case volumes, such as the same time unit and stratification. rRMSE was more useful than MAE for comparing scenarios with different case volumes, which allowed us to compare daily versus weekly and stratified versus unstratified estimates. The 95% PI coverage represents the percentage of estimates when the 95% PI included the final case count; the closer to 95%, the better the performance is. When the 95% PI coverage is near 100%, then PIs might be too wide to be informative.

We checked the dispersion ratio for the entire study period and for shorter periods of 14- and 21-day duration ending on Tuesdays to reflect the window lengths used in real time for diagnosis and onset. This was done using Poisson regression models of counts by each respective date to confirm whether a negative binomial data distribution was appropriate for this dataset.

Ethical Considerations

The New York City Health Department's institutional review board reviewed this work and determined it to be exempt human participants research under 45 CFR §46.104(d)(4)(ii) and (iii) (IRB No. 22 - 097). Analyses were performed using R version 4.2 and NobBS version 0.1.0. The frozen analytic line list, evaluation code, and codebook are available on GitHub [33].

XSL•FO
RenderX

Results

Data Lags and Interview Completeness

Among 3305 New York City residents diagnosed with mpox from July 8 through September 30, 2022, the median lag in days from diagnosis to diagnosis report was 4 (IQR 3-5, 90th percentile: 6). Lags decreased as the epidemic progressed, from a median lag of 4 days for patients diagnosed in July to 3 days for those diagnosed in September (Table 2). Of 3305 patients diagnosed with mpox, 2558 (77%) were probable cases, with a median lag in days from diagnosis to diagnosis report of 4 (IQR 3 - 5, 90th percentile: 6). The remaining 747 (23%) were confirmed cases, with a shorter median lag of 3 (IQR 2 - 4, 90th percentile: 5) days (Table 2). Of the 3305 diagnosed patients, 2429 (73%) had a fully or partially completed interview (Table S1 in Multimedia Appendix 1). Typically, the interview was conducted within a median of 1 (IQR 1 - 4) day of when the Health Department was notified of a confirmed or probable case, and a median of 10 (IQR 7 - 14) days of disease onset. The interview success rate was steady by diagnosis week, with a weekly median of 73% of patients successfully interviewed (range 64% - 80%).

XSL•FO
RenderX

Table . Lags from diagnosis to diagnosis report and from onset to onset report among New York City residents with confirmed or probable mpox diagnosis or onset from July 8 through September 30, 2022, by case status and month.

Date of interest, period, and stratification	Median number of days from date of interest to report of date of interest (IQR), 90th percentile			Values, n			*P* value[a] for Kruskal-Wallis test across race or ethnicity strata		
	Confirmed + Probable	Confirmed	Probable	Confirmed + Probable	Confirmed	Probable	Confirmed + Probable	Confirmed	Probable
Diagnosis									
July 8-September 30							.75	.47	.54
Unstratified	4 (3-5), 6	3 (2-4), 5	4 (3-5), 6	3305	747	2558	—[b]	—	—
Black or African American	4 (3-5), 6	3 (2-4), 5	4 (3-5), 6	919	231	688			
Hispanic or Latino	4 (3-5), 6	3 (2-4), 5	4 (3-5), 6	1131	257	874			
White	4 (3-5), 6	3 (3-4), 5	4 (3-5), 6	716	118	598			
July 8 - 31							.33	.34	.39
Unstratified	4 (3-5), 6	4 (3-4), 5	4 (3-5), 6	1458	101	1357	—	—	—
Black or African American	4 (3-5), 6	4 (3-5), 5	4 (3-5), 6	388	29	359			
Hispanic or Latino	4 (3-5), 6	3 (3-4), 5	4 (3-5), 6	448	27	421			
White	4 (3-5), 6	4 (3-5), 5	4 (3-5), 6	395	22	373			
August 1 - 31							.93	>.99	.76
Unstratified	4 (3-5), 7	3 (2-4), 5	4 (3-6), 7	1463	472	991	—	—	—
Black or African American	4 (3-5), 7	3 (2-5), 5	4 (3-6), 7	421	153	268			
Hispanic or Latino	4 (3-5), 6	3 (2-4), 6	4 (3-6), 6	528	157	371			
White	4 (3-5), 7	3 (2-4), 5	4 (3-6), 8	262	72	190			
September 1 - 30							.60	.35	.98
Unstratified	3 (2-3), 4	3 (1-3), 4	3 (2-4), 4	384	174	210	—	—	—
Black or African American	3 (2-3), 4	3 (1-3), 4	3 (2-4), 4	110	49	61			
Hispanic or Latino	3 (2-3), 4	3 (1-3), 4	3 (2-3), 4	155	73	82			
White	3 (2-4), 5	3 (2-4), 4	3 (2-4), 5	59	24	35			
Onset									
July 8-September 30							.97	.74	.97
Unstratified	10 (7 - 14), 18	9 (6-13), 18	10 (8 - 14), 18	2278	542	1736	—	—	—
Black or African American	10 (7 - 14), 18	9 (6-12), 17	10 (8 - 14), 18	648	174	474			

XSL•FO
RenderX

Date of interest, period, and stratification		Median number of days from date of interest to report of date of interest (IQR), 90th percentile			Values, n			*P* value[a] for Kruskal-Wallis test across race or ethnicity strata		
		Confirmed + Probable	Confirmed	Probable	Confirmed + Probable	Confirmed	Probable	Confirmed + Probable	Confirmed	Probable
	Hispanic or Latino	10 (7 - 14), 18	8 (7-13), 20	10 (7 - 14), 18	876	215	661			
	White	10 (8 - 13), 17	9 (6-12), 17	10 (8 - 13), 18	501	94	407			
July 8 - 31								.21	.66	.09
	Unstratified	11 (9 - 15), 19	11 (7 - 17), 21	11 (9 - 14), 19	1185	100	1085	—	—	—
	Black or African American	11 (9 - 15), 19	11 (7 - 14), 18	11 (9 - 15), 19	324	28	296			
	Hispanic or Latino	11 (9 - 15), 20	9 (7-17), 21	11 (9 - 15), 20	412	35	377			
	White	11 (9 - 14), 18	11 (9 - 16), 29	11 (9 - 13), 18	311	21	290			
August 1 - 31								.72	.30	.93
	Unstratified	9 (6-12), 16	8 (6-12), 18	9 (6-12), 15	866	338	528	—	—	—
	Black or African American	9 (6-12), 16	9 (6-13), 18	9 (6-12), 15	254	116	138			
	Hispanic or Latino	9 (6-12), 16	9 (7-13), 20	9 (6-12), 16	363	132	231			
	White	9 (6-12), 14	8 (5-11), 16	9 (7-12), 14	158	59	99			
September 1 - 30								.79	.87	.74
	Unstratified	8 (6-10), 14	8 (5-10), 14	8 (6-11), 14	227	104	123	—	—	—
	Black or African American	8 (5-10), 14	8 (5-11), 14	7 (5-10), 15	70	30	40			
	Hispanic or Latino	8 (6-10), 14	8 (6-10), 14	8 (6-11), 14	101	48	53			
	White	8 (6-11), 14	8 (5-9), 13	8 (6-12), 20	32	14	18			

[a]*P* values were unadjusted for multiple comparisons.

[b]Not applicable. Em dashes indicate there was no statistical test performed for unstratified values.

Of patients who were not interviewed, 88% (n=767) had missing onset dates and 28% (n=248) had unknown race or ethnicity (Table S1 in Multimedia Appendix 1). Race or ethnicity distributions were similar between patients who were and were not interviewed, except 39% (n=943) of interviewed patients were Hispanic or Latino, compared with only 21% (n=188) of not interviewed patients (Table S1 in Multimedia Appendix 1). The lower interview success rate among Hispanic or Latino patients could have reduced hindcasting performance for this stratum.

Separately, during the study period, 2278 patients had a recorded mpox illness onset date, and the median lag in days from onset to onset report was 10 (IQR 7 - 14, 90th percentile: 18). Lags decreased from a median of 11 days for patients with onset in July to 8 days in September (Table 2). Of 2278 patients with an illness onset date, 1736 (76%) were probable cases, with a median lag from onset to onset report of 10 (IQR 8 - 14, 90th percentile: 18) days. The remaining 542 (24%) were confirmed cases, with a shorter median lag of 9 (IQR 6 - 13, 90th percentile: 18) days (Table 2).

There was no statistically significant difference at α=.05 across race or ethnicity groups in the lag from diagnosis to diagnosis report or the lag from onset to onset report, overall or in any individual month based on the results of Kruskal-Wallis tests

XSL•FO
RenderX

(Table 2). Of 2278 patients with an onset date, 53 (2%) purportedly had onset after diagnosis, representing recall or data entry quality issues. Of the remaining 2225, the median lag in days from onset to diagnosis was 4 (IQR 2 - 7, 90th percentile: 10) (Table S2 in Multimedia Appendix 1).

Of 3305 patients diagnosed during this period, 812 (25%) were missing onset date (Table 3). Of these, 230 (28%) also had unknown race or ethnicity (Table 3). Onset date missingness increased with time, from 19% (n=278) for patients diagnosed in July to 31% (n=117) for those diagnosed in September (Table 4).

Counts of cases for the full study period and for 14-day windows by diagnosis date were consistently overdispersed in Poisson regression models by diagnosis date and less so for 21-day windows by onset date (Table S3 in Multimedia Appendix 1), supporting use in NobBS of a negative binomial case distribution.

Table . New York City residents diagnosed with mpox from July 8 through September 30, 2022, by onset date missingness, race or ethnicity, and interview status.

Patient characteristic	Missing onset date (n=812), n (column %)	Total (n=3305), n (column %)
Race or ethnicity		
Asian or Pacific Islander	20 (2.5)	109 (3.3)
Black or African American	221 (27.2)	919 (27.8)
Hispanic or Latino	182 (22.4)	1131 (34.2)
White	151 (18.6)	716 (21.7)
Other	8 (1.0)	56 (1.7)
Unknown	230 (28.3)	374 (11.3)
Interviewed		
Yes	45 (5.5)	2429 (73.5)
No	767 (94.5)	876 (26.5)

Table . New York City residents diagnosed with mpox from July 8 through September 30, 2022, by onset date missingness and diagnosis month.

Diagnosis month	Missing onset date, n (row %)	Total, n
July	278 (19.1)	1458
August	417 (28.5)	1463
September	117 (30.5)	384
Total	812 (24.6)	3305

Scenario Performance

Moving Window Lengths

For daily hindcasting unstratified by race or ethnicity, both by diagnosis and onset date, no single scenario performed best across MAE, rRMSE, 95% PI coverage, and average score (Table 5, Table S4 in Multimedia Appendix 1). For hindcasting by diagnosis date, as moving window lengths increased, the average score generally improved (increased), MAE generally improved (decreased), and rRMSE worsened (increased). Patterns were inconsistent for hindcasting by onset date.

For hindcasting by diagnosis date, the average score for the 14-day moving window used in real time was 0.27, with other scenarios ranging from 0.27 to 0.47 (Table 5). The MAE for the 14-day moving window was 9, with other scenarios ranging from 3 to 9. The rRMSE for the 14-day window was 0.23, with other scenarios ranging from 0.25 to 0.30. The 95% PI coverage for the 14-day window was 96%, with other scenarios ranging from 93% to 100%.

XSL•FO
RenderX

Table . Performance measures for diagnosis date–based hindcasting approaches in Nowcasting by Bayesian Smoothing, applied to daily case counts of New York City residents with mpox diagnosis from July 13 through September 27, 2022 (metrics calculated on last 7 days of hindcast, excluding weekends).

Stratification and scenario number	Window length (days)[a], n	Mean absolute error	Relative root mean square error	Number of estimates when the 95% prediction interval included the final case count (95% prediction interval coverage)	Number of estimates evaluated (number of models run)	Average score
Unstratified						
1[b]	14	9.04	0.23	48 (96.00)	50 (10)	0.27
2	21	8.73	0.25	42 (93.33)	45 (9)	0.28
3	28	7.18	0.25	37 (92.50)	40 (8)	0.27
4	35	5.09	0.27	35 (100.00)	35 (7)	0.41
5	42	3.93	0.29	29 (96.67)	30 (6)	0.44
6	49	2.88	0.30	24 (96.00)	25 (5)	0.47
Black or African American						
7[b]	14	2.90	0.30	49 (98.00)	50 (10)	0.39
8	21	2.16	0.32	44 (97.78)	45 (9)	0.41
9	28	1.77	0.33	39 (97.50)	40 (8)	0.41
10	35	1.43	0.37	35 (100.00)	35 (7)	0.48
11	42	1.10	0.41	30 (100.00)	30 (6)	0.49
12	49	1.24	0.51	25 (100.00)	25 (5)	0.49
Hispanic or Latino						
13[b]	14	3.42	0.34	46 (92.00)	50 (10)	0.32
14	21	3.09	0.35	42 (93.33)	45 (9)	0.33
15	28	2.70	0.39	37 (92.50)	40 (8)	0.33
16	35	1.69	0.37	35 (100.00)	35 (7)	0.49
17	42	1.60	0.48	30 (100.00)	30 (6)	0.50
18	49	1.28	0.50	25 (100.00)	25 (5)	0.52
White						
19[b]	14	2.10	0.32	48 (96.00)	50 (10)	0.45
20	21	1.69	0.39	44 (97.78)	45 (9)	0.52
21	28	1.38	0.41	39 (97.50)	40 (8)	0.54
22	35	1.20	0.46	35 (100.00)	35 (7)	0.64
23	42	1.10	0.49	30 (100.00)	30 (6)	0.70
24	49	0.76	0.49	25 (100.00)	25 (5)	0.75
All stratified						
25[b]	14	2.81	0.32	143 (95.33)	150 (10)	0.38
26	21	2.31	0.35	130 (96.30)	135 (9)	0.41
27	28	1.95	0.38	115 (95.83)	120 (8)	0.42
28	35	1.44	0.40	105 (100.00)	105 (7)	0.53
29	42	1.27	0.46	90 (100.00)	90 (6)	0.55
30	49	1.09	0.50	75 (100.00)	75 (5)	0.57

[a]14-, 21-, 28-, 35-, 42-, and 49-day nowcasts started on July 26, August 2, August 9, August 16, August 23, and August 30, 2022, respectively, to

XSL•FO
RenderX

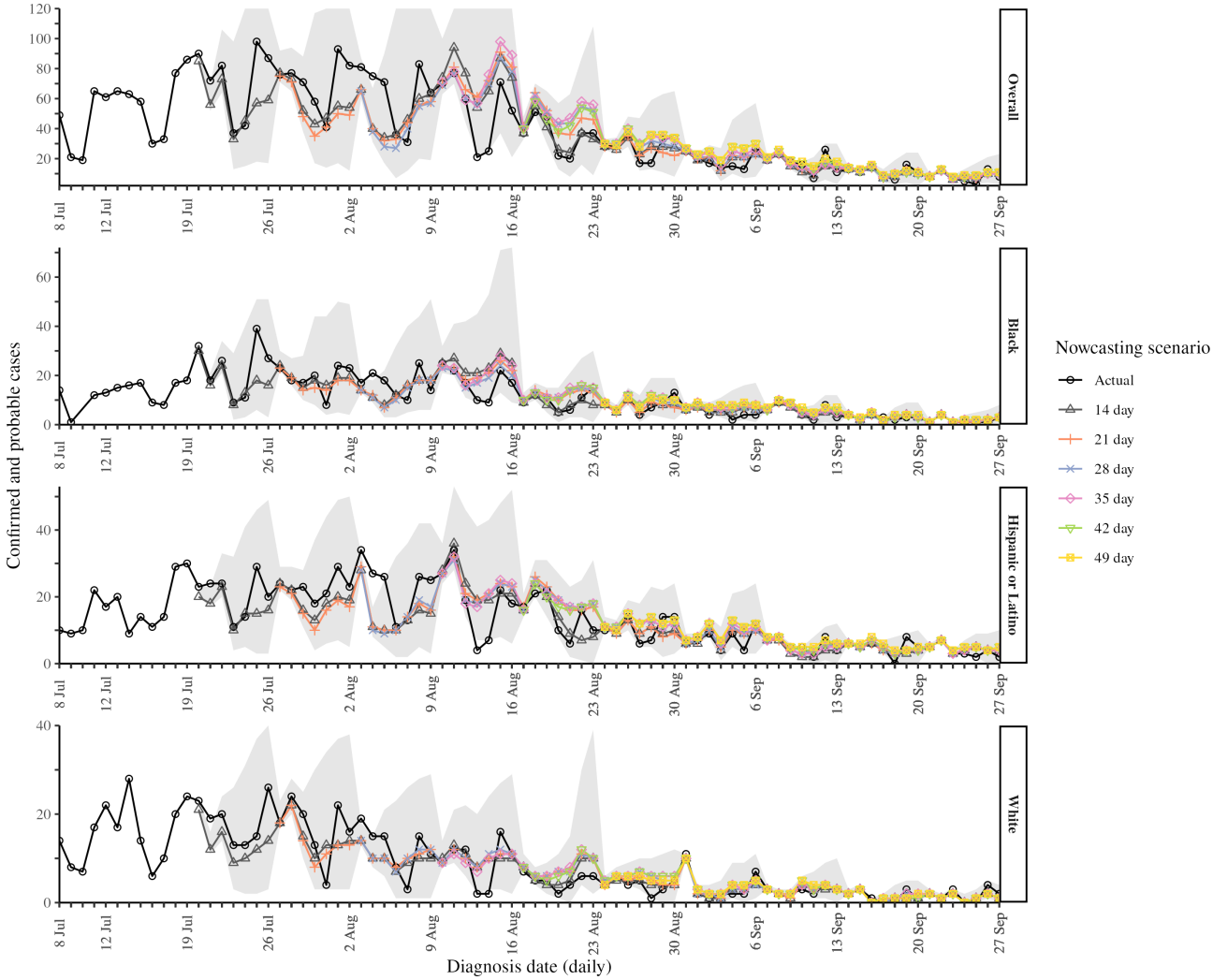
provide 2, 3, 4, 5, 6, or 7 weeks of Wednesday-Tuesday data since study start date July 8, 2022. We mimicked nowcasts weekly, ending September 27, 2022, as the last Tuesday during the study period.

[b]Indicates scenario applied in real time at the New York City Health Department.

For hindcasting by onset date, the average score for the 21-day moving window used in real time was 0.23, with other scenarios ranging from 0.16 for the 42-day window to 0.30 for the 14-day window (Table S4 in Multimedia Appendix 1). The MAE for the 21-day moving window was 12, with other scenarios ranging from 7 to 11. The rRMSE for the 21-day window was 1.07, with other scenarios ranging from 0.75 to 1.42. The 95% PI coverage for the 21-day window was 84%, with other windows ranging from 75% to 99% (Table S4 in Multimedia Appendix 1).

Overall, hindcasts underestimated diagnoses in early August 2022, on the downslope of the epidemic curve, then overestimated diagnoses in late August (Figure 2 and Figure S1 in Multimedia Appendix 1). Hindcasting overestimated onsets throughout the study period, except for the 14-day daily and 2-week weekly moving windows, which underestimated cases at points in early and late August 2022 (Figures S2 and S3 in Multimedia Appendix 1). Lags from onset to onset report decreased rapidly in July and August (Figure S4 in Multimedia Appendix 1); the shortening delay distribution over time might have led NobBS to overestimate onsets. By September 2022, diagnoses and onsets were low and stable, and both daily and weekly hindcast estimates, regardless of window length, were close to final diagnosis counts (Figure 2 and Figures S1-S3 in Multimedia Appendix 1).

Figure 2. Comparison of 7-day hindcasts conducted on Wednesdays using various moving window lengths at the daily time unit for confirmed and probable mpox cases among New York City residents diagnosed from July 8 through September 27, 2022, overall and stratified by 3 race or ethnicity groups. Final case counts reported as of September 1, 2023, are shown in black. The 95% prediction interval is shown in gray for the 14-day window, which was the scenario implemented in real time. The y-axis for overall diagnoses was truncated at 120 for clarity, but the observed upper bound of the 95% prediction interval for the 14-day window was 252 on August 16, 2022.



Stratification

For daily diagnosis hindcasts stratified by race or ethnicity, the average score for the 14-day moving window used in real time was 0.38, with other scenarios ranging from 0.41 to 0.57 (Table 5). The average score was higher in stratified estimates compared with unstratified estimates. When evaluating race or ethnicity strata individually, hindcasts for White patients had the highest performance (higher average scores ranging from 0.45-0.75), while hindcasts for Black or African American and

XSL•FO
RenderX

Hispanic or Latino patients had lower performance (ranging from 0.39-0.49 and 0.32-0.52, respectively). Worse performance in particular strata could be explained by sparser counts and epidemic trends that are difficult to estimate or by minor differences in the delay distribution and interview success rates across strata.

The rRMSE for the 14-day moving window was 0.32, with other scenarios ranging from 0.35 to 0.50 (Table 5). The 95% PI coverage for the stratified 14-day diagnosis window was 95%, with other scenarios ranging from 96% to 100%. For stratified daily onset hindcasts, the average score for the 21-day window used in real time was 0.36, with other scenarios ranging from 0.36 to 0.54. The rRMSE for the 21-day window was 1.22; others ranged from 0.91 to 1.71 (Table S4 in Multimedia Appendix 1). The 95% PI coverage for the stratified 21-day onset window was 95%; others ranged from 89% to 97%. For any given moving window length, rRMSE increased (worsened) for stratified compared with unstratified estimates in both diagnosis and onset-based hindcasts. For any given moving window length, the 95% PI coverage was not consistently closer to 95% in either the stratified or unstratified scenario.

Weekly Time Unit

For unstratified weekly diagnosis hindcasts, the average score remained stable at different window lengths, ranging from 0.25 to 0.30 (Table S5 in Multimedia Appendix 1). This was comparable to the performance of unstratified daily diagnosis hindcasts in shorter window lengths (14, 21, and 28 days) and worse in longer window lengths (35, 42, and 49 days; Table 5). The rRMSE for unstratified weekly diagnosis hindcasts ranged from 0.21 to 0.37 (Table S5 in Multimedia Appendix 1). This was similar to the rRMSE for daily unstratified diagnosis hindcasts, which ranged from 0.23 through 0.30 across moving windows (Table 5). The 95% PI coverage ranged from 83% to 100% (Table S5 in Multimedia Appendix 1).

For unstratified weekly onset hindcasts, the average score was poor across all moving window lengths, ranging from 0.09 to 0.18 (Table S5 in Multimedia Appendix 1), and was worse than the average scores at daily resolution (Table S4 in Multimedia Appendix 1). The rRMSE ranged from 0.24 to 1.10 (Table S5 in Multimedia Appendix 1). This was similar to rRMSE in unstratified daily onset hindcasts, which ranged from 0.75 to 1.42 (Table S4 in Multimedia Appendix 1). The 95% PI coverage ranged from 60% to 100% (Table S5 in Multimedia Appendix 1). For a given moving window length, rRMSE typically increased (worsened) weekly compared with daily diagnosis hindcasts but decreased (improved) weekly compared with daily onset hindcasts. Weekly hindcasts generally had worse 95% PI coverage than their daily counterpart. The lowest performing window length based on 95% PI coverage was much worse for unstratified weekly scenarios (83% for diagnosis and 60% for onset; Table S5 in Multimedia Appendix 1) than for daily scenarios (93% for diagnosis and 75% for onset; Table 5, Table S4 in Multimedia Appendix 1).

Discussion

Principal Findings

In evaluating NobBS for the 2022 mpox outbreak in New York City, we faced challenges in developing input files using the onset date. In addition, no moving window length consistently performed best. Daily time units performed better than weekly, and stratifying by race or ethnicity improved performance.

A key challenge in developing input files was that the onset date was frequently missing, which is a common challenge for mpox data collected via patient interviews [34]. When the onset date was available, it was usually after a long delay; the 90th percentile of delay from onset to onset report was 18 days (Table 2), reducing the usefulness of shorter moving window lengths. Furthermore, the onset report date was not a standardized field in our disease surveillance database, which led to implementation delays during the public health emergency. Performance was better when hindcasting by diagnosis date than by onset date, as expected given shorter lags from diagnosis to diagnosis report than from onset to onset report and missingness in onset date.

The choice of moving window length and whether to stratify by race or ethnicity had less influence on hindcasting performance than the choice of aggregating to daily or weekly time units. We had anticipated that with sparsity from relatively few cases in this outbreak, nowcasting at weekly aggregation might improve performance. This was not borne out, possibly because of greater difficulty in estimating the epidemic trend using fewer data points. Hindcasting was more accurate when counts were low and stable, toward the end of the outbreak. Others have also found that forecasting performance metrics varied between early and declining mpox outbreak phases [32]. This underscores the need for nowcasting methods that will reliably perform well as epidemics grow, peak, and decline.

Stratifying by race or ethnicity improved performance, and the highest average scores were observed for White patients. Performance at shorter windows was lowest for hindcasts of Hispanic or Latino patients, possibly due to a lower interview success rate.

Limitations

Several data quality limitations were noted during project implementation. First, a quarter of diagnosed patients had missing onset dates, which made onset dates less reliable than diagnosis dates for monitoring trends. Patient interviews were the primary source for the onset date. Some patients may have refused interviews due to the sensitive nature of revealing a sexual history in the context of their mpox diagnosis. Generally, surveys about sexual history have participant refusal rates of 25% - 35% [35]. Another reason for missingness is that some patients could not recall their onset date.

As onset dates and race and ethnicity data were often collected during interviews, the stratified and onset-based nowcasts relied on incomplete reports (Table 3, Table S1 in Multimedia Appendix 1). Unstratified, diagnosis-based hindcasts were the only type of hindcast evaluated that relied only on complete and timely laboratory reporting data. Additionally, the median

XSL•FO
RenderX

delay from onset to report decreased rapidly from the study start until late August (Figure S4 in Multimedia Appendix 1). Shortening delay distributions could have led NobBS to overestimate onsets in August. Shorter moving windows started with input data from the peak and early decline of the outbreak, while delay distributions and epidemic trends were rapidly changing. Longer moving windows, which appeared to be associated with better average scores, only began once case counts had stabilized, limiting our ability to directly compare window lengths.

Additionally, we included both confirmed and probable cases. Delays for both diagnosis to diagnosis report and onset to onset report were slightly shorter for confirmed than probable cases. While differences in delays by case status were minor, accounting for case status might improve accuracy. Additionally, stratified estimates were limited to Black or African American, Hispanic or Latino, and White patients, while unstratified estimates were for all patients, regardless of race or ethnicity, reducing our ability to directly compare stratified and unstratified estimates.

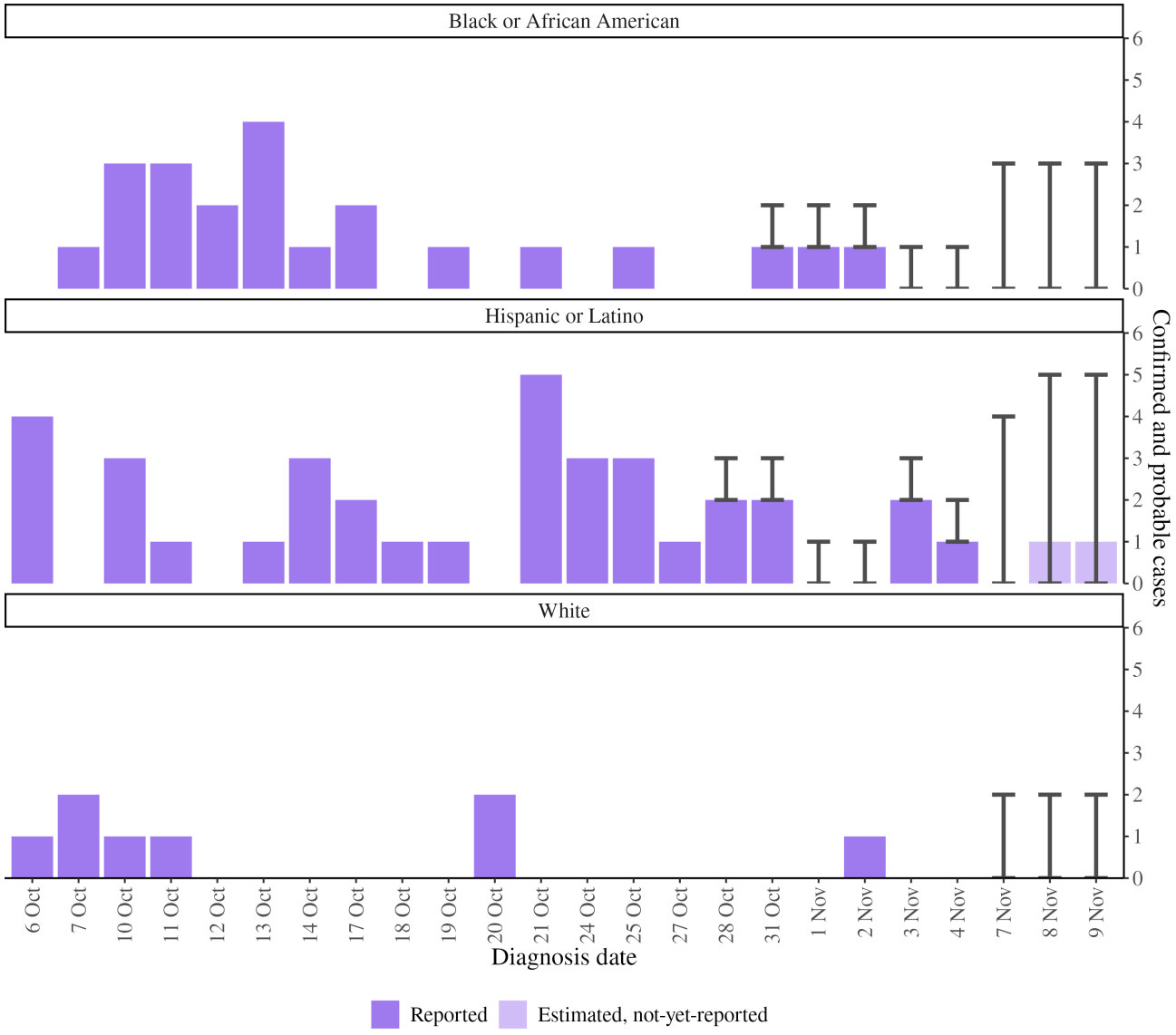
Although NobBS accounts for reporting delays, it does not account for other limitations of reportable disease data, including underascertainment, underreporting, and misdiagnosis or misclassification [19]. NobBS also does not account for external determinants influencing epidemic trends, such as behavioral changes or public health interventions. Our study period began after commercial laboratory testing became available, which nearly coincided with the epidemic peak, so we were unable to evaluate nowcasting performance during initial epidemic growth. We observed trade-offs in evaluation metrics, for example, scenarios of improved PI coverage with decreased accuracy (Table 5, Tables S4 and S5 in Multimedia Appendix 1), which could be related to overfit models or overconfident PIs. Additionally, the maximum delay used in NobBS of the moving window length minus 1 meant that window lengths were longer than the 90th percentile of observed delays for almost all moving windows. This could explain why changing the window lengths did not have a major impact on performance. Also, lags from diagnosis to report were almost universally less than 1 week, and nowcasting at weekly resolution may not be warranted for such short reporting delays. We did not compare NobBS with other nowcasting methods, such as generalized additive models [34,36], nor did we assess methods developed for the purpose of estimating the time-varying effective reproduction number instead of observed case counts [31].

Practice Implications

Accurate nowcasts can facilitate real-time trend monitoring and reporting to policymakers. Stratifying nowcasts by key demographic characteristics associated with inequities, including disaggregated race or ethnicity groups, can help public health authorities quickly identify and remediate inequities faster than monitoring epidemic curves, without accounting for data lags. For example, on November 10, 2022, in the context of declining overall case counts and a focus on ensuring equitable access to interventions, we presented stratified nowcasting results to the Incident Command System leadership, highlighting that the number of recent estimated cases, even with uncertainty, was disproportionately higher among Hispanic or Latino New Yorkers (Figure 3). This finding was borne out after data fully accrued (Figure 4); final daily case counts were within the narrow 95% PIs for estimated case counts.

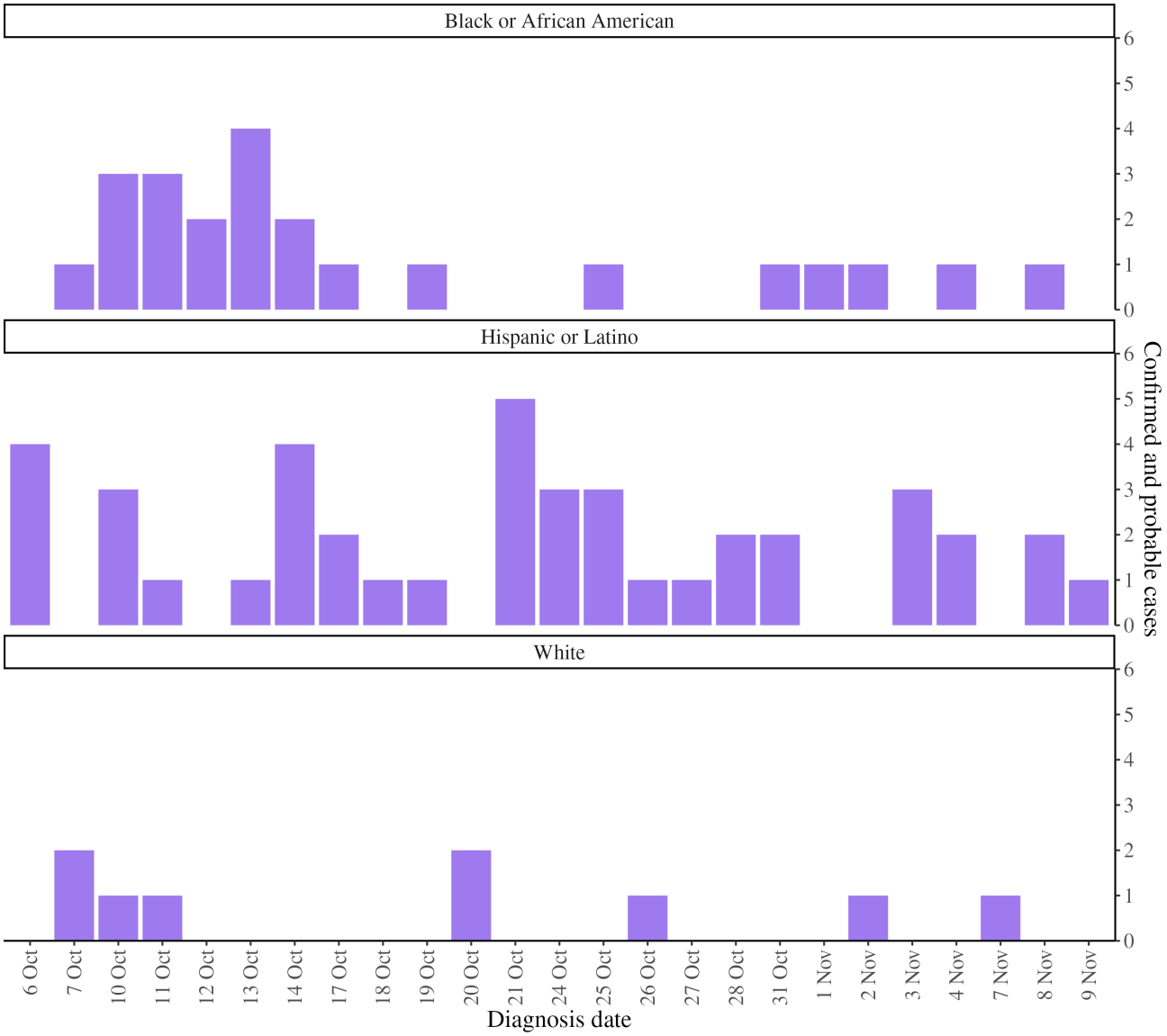
XSL•FO
RenderX

Figure 3. Hindcast visualization of reported and estimated (not-yet-reported) mpox cases diagnosed among New York City residents, presented to Incident Command System leadership on November 10, 2022. The error bars represent 95% prediction intervals.



XSL•FO
RenderX

Figure 4. Mpox cases diagnosed among New York City residents for the same period as Figure 3, after data fully accrued.



We recommend stratifying nowcasts to monitor differences in epidemic trends across patient groups and to improve performance, as well as using diagnosis date rather than onset date. For future outbreaks, health departments can strengthen preparedness to rapidly initiate nowcasting during public health emergencies by populating a field for onset report date directly in the surveillance database. Imputing the onset date might be necessary to improve completeness [31].

Performance metrics were sensitive to NobBS implementation details, and no single moving window length emerged as best performing. Health departments need reliable tools to initiate daily nowcasting by diagnosis date within the first few weeks of a public health emergency, to conduct interim performance evaluations to assess accuracy, and to pinpoint which adjustments to make to improve performance while emergencies are ongoing. Tools such as the scoringutils R package [37] could facilitate rapid evaluations and adjustments. Additional practical guidance is needed for health departments on how to optimize nowcasting, including how to add robustness by using multiple distinct methods, and how to best evaluate performance.

Acknowledgments

The authors thank the New York City Health Department's Incident Command System staff who worked on the mpox response, particularly the Surveillance and Investigations Unit for conducting patient interviews. The authors also thank Naama Kipperman for contributing to data extraction and Chasmandeep Bring for administering the Health Department's R server. They thank Rebecca Kahn for developing the code for a prior COVID-19 nowcasting evaluation [24], which served as the basis for this evalution. They thank Sarah McGough for providing code and guidance to calculate the log score using Nowcasting by Bayesian Smoothing output.

XSL•FO
RenderX

Data Availability

The line list of cases analyzed in this evaluation, the evaluation code, and the codebook are available in the mpox_nowcast_eval repository [33].

Authors' Contributions

RR led data extraction, cleaning, analysis, and results interpretation. AW reviewed code for data extraction, cleaning, and analysis. JB provided guidance on local datasets, surveillance workflow, and data point selection. NB and LEJ co-led a team conducting patient interviews, including collecting data on onset dates and race and ethnicity. RRO and AD led health equity efforts, including emphasizing the importance of stratifying by race or ethnicity when monitoring epidemic trends. SKG conceived this evaluation, provided oversight, applied lessons learned from a prior COVID-19 nowcasting evaluation, and interpreted results. RR and SKG drafted the paper. AW, JB, NB, RRO, AD, and LEJ critically reviewed the paper. All authors gave final approval of the submitted version. The authors did not use generative artificial intelligence for any portion of paper writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional details about patient characteristics by interview status, data lags, assessment of overdispersion in case counts, and Nowcasting by Bayesian Smoothing performance metrics by onset date and at weekly resolution.
[DOCX File, 842 KB - ojphi_v17i1e56495_app1.docx]

References

1. Laurenson-Schafer H, Sklenovská N, Hoxha A, et al. Description of the first global outbreak of mpox: an analysis of global surveillance data. Lancet Glob Health 2023 Jul;11(7):e1012-e1023. [doi: 10.1016/S2214-109X(23)00198-5] [Medline: 37349031]
2. Gnanaprakasam R, Keller M, Glassman R, et al. Mpox in the New York metropolitan area, summer 2022. J Med Virol 2023 Apr;95(4):e28699. [doi: 10.1002/jmv.28699] [Medline: 36951318]
3. 2022 health advisory 6: CDC advisory monkeypox virus infection in the United States and other non-endemic countries—2022. New York City Health Department. 2022. URL: https://www1.nyc.gov/assets/doh/downloads/pdf/han/advisory/2022/cdc-monkeypox-us.pdf [accessed 2022-10-31]
4. Health advisory: monkeypox cases not associated with travel to areas where monkeypox is enzootic. New York State Department of Health. 2022. URL: https://health.ny.gov/diseases/communicable/zoonoses/monkeypox/docs/2022-07-08_han.pdf [accessed 2022-11-30]
5. Emergency order no. 158, declaration of local state of emergency. City of New York Office of the Mayor. 2022. URL: https://www.nyc.gov/assets/home/downloads/pdf/executive-orders/2022/eeo-158.pdf [accessed 2023-07-11]
6. Determination that a public health emergency exists. U.S. Department of Health & Human Services. 2022. URL: https://aspr.hhs.gov/legal/PHE/Pages/monkeypox-4Aug22.aspx [accessed 2024-07-30]
7. New York City declares end to mpox outbreak after nation-leading response. New York City Health Department. 2023. URL: https://www.nyc.gov/site/doh/about/press/pr2023/nyc-declares-end-to-mpox-outbreak.page [accessed 2023-06-13]
8. Mpox. New York City Health Department. 2024. URL: https://www.nyc.gov/site/doh/health/health-topics/mpox.page [accessed 2024-01-29]
9. Mpox cases by age and gender and race and ethnicity. CDC. 2024. URL: https://archive.cdc.gov/www_cdc_gov/poxvirus/mpox/response/2022/demographics.html [accessed 2025-01-07]
10. Mpox 2022 summary. New York City Health Department. 2023. URL: https://www.nyc.gov/assets/doh/downloads/pdf/monkeypox/mpox-response-data-summary.pdf [accessed 2023-05-12]
11. Mpox virus infection 2022 case definition. CDC. 2023. URL: https://ndc.services.cdc.gov/case-definitions/monkeypox-virus-infection-2023/ [accessed 2025-01-07]
12. Blackburn D, Roth NM, Gold JAW, et al. Epidemiologic and clinical features of Mpox in transgender and gender-diverse adults - United States, May-November 2022. MMWR Morb Mortal Wkly Rep 2022 Dec 30;71(5152):1605-1609. [doi: 10.15585/mmwr.mm715152a1] [Medline: 36580418]
13. Kirby T. What happened to the mpox pandemic? Lancet 2023 Sep 16;402(10406):949-950. [doi: 10.1016/S0140-6736(23)01956-6] [Medline: 37717593]
14. Calling for a more equitable distribution of monkeypox vaccine. Harlem United. 2022. URL: https://www.harlemunited.org/access-to-monkeypox-vaccine/ [accessed 2024-01-24]
15. Du M, Sun H, Zhang S, et al. Global epidemiological features of human monkeypox cases and their associations with social-economic level and international travel arrivals: a systematic review and ecological study. Int J Public Health 2023;68:1605426. [doi: 10.3389/ijph.2023.1605426] [Medline: 36743344]

XSL•FO
RenderX

16. Lash MK, Latham NH, Chan PY, et al. Racial and socioeconomic equity of tecovirimat treatment during the 2022 mpox emergency, New York, New York, USA. Emerg Infect Dis 2023 Nov;29(11):2353-2357. [doi: 10.3201/eid2911.230814] [Medline: 37796277]
17. Risk communication and community engagement public health advice on understanding, preventing and addressing stigma and discrimination related to monkeypox. WHO. 2022. URL: https://cdn.who.int/media/docs/default-source/documents/emergencies/final-rcce-interim-guidance-on-using-inclusive-language-to-address-stigma-and-discrimination-1-september-2022.pdf [accessed 2025-01-07]
18. Eaton LA, Driffin DD, Kegler C, et al. The role of stigma and medical mistrust in the routine health care engagement of Black men who have sex with men. Am J Public Health 2015 Feb;105(2):e75-e82. [doi: 10.2105/AJPH.2014.302322] [Medline: 25521875]
19. Bragazzi NL, Woldegerima WA, Iyaniwura SA, et al. Knowing the unknown: the underestimation of monkeypox cases. Insights and implications from an integrative review of the literature. Front Microbiol 2022;13:1011049. [doi: 10.3389/fmicb.2022.1011049] [Medline: 36246252]
20. Guenot M. NYC doctor says the medical system is failing monkeypox victims and putting privileged patients first. Business Insider. 2022. URL: https://www.businessinsider.com/monkeypox-doctor-medical-system-failing-victims-new-york-2022-8 [accessed 2024-01-24]
21. Gilyard K. A health equity expert on making monkeypox messaging more inclusive. STAT. 2022. URL: https://www.statnews.com/2022/08/05/stella-safo-monkeypox-messaging-inclusive/ [accessed 2024-01-24]
22. Gaffney T. With support on monkeypox hard to come by, queer communities turn to one another. STAT. 2022. URL: https://www.statnews.com/2022/08/11/with-support-on-monkeypox-hard-to-come-by-queer-communities-turn-to-one-another/ [accessed 2024-01-26]
23. McGough SF, Johansson MA, Lipsitch M, Menzies NA. Nowcasting by Bayesian Smoothing: a flexible, generalizable model for real-time epidemic tracking. PLoS Comput Biol 2020 Apr;16(4):e1007735. [doi: 10.1371/journal.pcbi.1007735] [Medline: 32251464]
24. Greene SK, McGough SF, Culp GM, et al. Nowcasting for real-time COVID-19 tracking in New York City: an evaluation using reportable disease data from early in the pandemic. JMIR Public Health Surveill 2021 Jan 15;7(1):e25538. [doi: 10.2196/25538] [Medline: 33406053]
25. Menkir TF, Cox H, Poirier C, et al. A nowcasting framework for correcting for reporting delays in malaria surveillance. PLoS Comput Biol 2021 Nov;17(11):e1009570. [doi: 10.1371/journal.pcbi.1009570] [Medline: 34784353]
26. Rotejanaprasert C, Ekapirat N, Areechokchai D, Maude RJ. Bayesian spatiotemporal modeling with sliding windows to correct reporting delays for real-time dengue surveillance in Thailand. Int J Health Geogr 2020 Mar 3;19(1):4. [doi: 10.1186/s12942-020-00199-0] [Medline: 32126997]
27. McGough SF, Johansson MA, Lipsitch M, Menzies NA. NobBS: Nowcasting by Bayesian Smoothing, version 0.1.0. The Comprehensive R Archive Network. 2020. URL: https://CRAN.R-project.org/package=NobBS [accessed 2024-12-02]
28. Electronic clinical laboratory reporting system. New York State Department of Health. 2023. URL: https://www.health.ny.gov/professionals/reportable_diseases/eclrs/ [accessed 2024-01-29]
29. Reporting results from orthopoxvirus, non-variola orthopoxvirus, and mpox virus diagnostic testing. CDC. 2024. URL: https://www.cdc.gov/mpox/hcp/laboratories/reporting-test-results.html [accessed 2025-01-07]
30. Instructions for using the universal reporting form (URF). New York City Health Department. 2022. URL: https://www.nyc.gov/site/doh/providers/reporting-and-services/hcp-urf.page [accessed 2022-07-30]
31. Charniga K, Madewell ZJ, Masters NB, Asher J, Nakazawa Y, Spicknall IH. Nowcasting and forecasting the 2022 U.S. mpox outbreak: support for public health decision making and lessons learned. Epidemics 2024 Jun;47:100755. [doi: 10.1016/j.epidem.2024.100755] [Medline: 38452454]
32. Bleichrodt A, Dahal S, Maloney K, Casanova L, Luo R, Chowell G. Real-time forecasting the trajectory of monkeypox outbreaks at the national and global levels, July-October 2022. BMC Med 2023 Jan 16;21(1):19. [doi: 10.1186/s12916-022-02725-2] [Medline: 36647108]
33. Rohrer R. Mpox nowcasting evaluation [analytic line list, evaluation code, and codebook]. New York City Health Department GitHub. 2024. URL: https://github.com/nychealth/mpox_nowcast_eval [accessed 2024-12-02]
34. Overton CE, Abbott S, Christie R, et al. Nowcasting the 2022 mpox outbreak in England. PLoS Comput Biol 2023 Sep;19(9):e1011463. [doi: 10.1371/journal.pcbi.1011463] [Medline: 37721951]
35. Fenton KA, Johnson AM, McManus S, Erens B. Measuring sexual behaviour: methodological challenges in survey research. Sex Transm Infect 2001 Apr;77(2):84-92. [doi: 10.1136/sti.77.2.84] [Medline: 11287683]
36. Wood S. Mgcv: mixed GAM computation vehicle with automatic smoothness estimation. The Comprehensive R Archive Network. 2023. URL: https://CRAN.R-project.org/package=mgcv [accessed 2024-12-02]
37. Bosse NI, Gruson H, Cori A, Leeuwen E, Funk S, Abbott S. Evaluating forecasts with scoringutils in R. arXiv. Preprint posted online on May 14, 2022. [doi: 10.48550/arXiv.2205.07090]

XSL•FO
RenderX

Abbreviations

MAE: mean absolute error
NobBS: Nowcasting by Bayesian Smoothing
PI: prediction interval
rRMSE: relative root mean square error

Edited by A Mavragani; submitted 31.01.24; peer-reviewed by A Bleichrodt, C Zhao, MS Torres Silva, V Lopez; revised version received 13.09.24; accepted 19.09.24; published 14.01.25.

Please cite as:
Rohrer R, Wilson A, Baumgartner J, Burton N, Ortiz RR, Dorsinville A, Jones LE, Greene SK
Nowcasting to Monitor Real-Time Mpox Trends During the 2022 Outbreak in New York City: Evaluation Using Reportable Disease Data Stratified by Race or Ethnicity
Online J Public Health Inform 2025;17:e56495
URL: https://ojphi.jmir.org/2025/1/e56495
doi:10.2196/56495

© Rebecca Rohrer, Allegra Wilson, Jennifer Baumgartner, Nicole Burton, Ray R Ortiz, Alan Dorsinville, Lucretia E Jones, Sharon K Greene. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 14.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of a Sentinel Hypertension Surveillance System in Mojo, East Shewa Zone, Oromia, Ethiopia: Concurrently Embedded Mixed Design Study

Abiyie Demelash Gashe[1], BSc, MPH; Yeshiwas Ayale Ferede[2], BSc, MPH; Dawit Zenebe Weldemichael[3], BSc, MPH; Aman Yesuf Endries[4], BSc, MPH

[1]Ethiopian Field Epidemiology Training Program, Department of Epidemiology, School of Public Health, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia

[2]Department of Reproductive Health, Teda Health Science College, Gondar, Ethiopia

[3]Department of Epidemiology, College of Health Science, Mekelle University, Mek'ele, Ethiopia

[4]Department of Epidemiology, School of Public Health, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia

Corresponding Author:
Abiyie Demelash Gashe, BSc, MPH
Ethiopian Field Epidemiology Training Program, Department of Epidemiology, School of Public Health, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia

Abstract

Background: In response to the increasing incidence and prevalence of hypertension, Ethiopia has been piloting hypertension control at the primary health care level in selected sentinel sites. However, no evaluation has been conducted and its success and failures have not been ascertained.

Objective: This study aimed to evaluate on whether sentinel hypertension surveillance system in Mojo City were operating efficiently and effectively.

Methods: A concurrently embedded mixed design (quantitative or qualitative) study was conducted in 2 sentinel health centers in Mojo city, Oromia region of Ethiopia. The usefulness and 9 system attributes were assessed via key informant interviews, observations, and record reviews. The qualitative data were analyzed manually via thematic analysis, whereas quantitative data were analyzed via SPSS Software version 25.0 (IBM Corp).

Results: The study invited 14 key informants, and all were willing to participate in the interview. The completeness and timeliness of reports were 98% and 100%, respectively. The sensitivity, positive predictive value, and representativeness were 45.3%, 92.6%, and 22%, respectively. Nearly three-fourths (10/14, 71%) of key informants perceived the system as flexible, while half thought it as unstable due to factors such as inadequate training and lack of supportive supervision and feedback system. Health facilities did not conduct routine data analysis and interpretation, nor did they use for action.

Conclusions: The surveillance system in Mojo city was simple, flexible, acceptable, and predictive but less sensitive, unrepresentative, and unstable. There is a need for implementing routine data analysis and use for action, adequate training, and feedback system for optimizing the system's performance and to ensure its sustainability.

(Online J Public Health Inform 2025;17:e72909) doi:10.2196/72909

KEYWORDS

hypertension; sentinel surveillance; public health surveillance; noncommunicable disease surveillance; Ethiopia

Introduction

Hypertension is the leading risk factor for cardiovascular diseases [1,2]. Approximately 1.28 billion adults aged 30 to 79 years have hypertension worldwide, with low- and middle-income countries bearing two-thirds of the global burden. Alarmingly, 46% of adults with hypertension are unaware of their condition while only 42% received appropriate diagnosis and treatment. Furthermore, 1 in 5 individuals with hypertension have successfully controlled their blood pressure (BP) [3]

While global trends in hypertension have shown significant increases each year, the rise is disproportionately seen in low- and middle-income countries especially in the African region [4,5]. By the end of 2021, the World Health Organization (WHO) of Africa exhibited the highest prevalence of hypertension (27%) compared with other regions [3]. In sub-Saharan Africa, the prevalence has been reported as high as 48% [5].

XSL•FO
RenderX

Hypertension is among one of the rising noncommunicable diseases (NCDs) [6-8] due to factors such as rapid urbanization, aging population as result of improved life expectancy, globalization of marketing and trade, and exposure to the social determinants of health [9,10]. The WHO estimated that the prevalence of hypertension in Ethiopia exceeds 20% and is alarmingly increasing over time [11]. Data from meta-analyses indicate that the prevalence of hypertension ranges from 9.3% to 30.3% in population-based studies, 7% to 37% in institution-based studies, and 13.2%% to 18.8% in hospital-based studies [12]. A substantial gap also exists in hypertension control in the country, with several hospital-based studies reporting BP control rates ranging from 26.4% to 52.5% [13-16].

Due to the rising incidence and prevalence of NCDs including hypertension, strengthening of the NCD surveillance system has become increasingly important. The WHO has emphasized the need for improved NCD surveillance systems to effectively tackle NCDs [17]. Consequently, Ethiopia has made a significant shift in its approach to public health surveillance. For over 3 decades, the country implemented surveillance of epidemic-prone communicable diseases under public health emergency management system [18]. However, country adopted its first NCD-inclusive public health emergency management guideline that included hypertension and diabetes mellitus by 2021 [19].

In addition, to enhance hypertension prevention and control, the Federal Ministry of Health, in collaboration with the WHO and Resolve to Save Lives, initiated a hypertension control project at the primary health care level including 74 health facilities and 200 health posts since 2019 [20]. These facilities have been used as hypertension sentinel sites to facilitate early detection, timely response, and data-driven decision making.

The target population for this surveillance system comprises adults aged 30 years and older. The major milestones of this initiative include screening 30% of the eligible population and achieving a 50% BP control. However, no evaluation has been conducted at either national or subnational levels. Hence, evaluating this program is crucial to ascertain its success and failures, share lessons learnt, guide evidence-based decision making, inform clinical practice, and improve overall hypertension control. Therefore, this study aimed to evaluate on whether the sentinel hypertension surveillance systems in 2 health centers in Mojo City, Oromia region of Ethiopia are operating effectively and efficiently.

Methods

Study Area

Mojo City is located in the East Shewa zone, Oromia region of Ethiopia at 74 km distance from Addis Ababa, the capital city of Ethiopia. Based on Ethiopia's 2007 census data, the estimated population of Mojo City was 58,941, of which 51% (30,060) were females. Adults aged 30 years and older constitute 29% (17,093/58,941) of the total population. Mojo City has 1 primary hospital, 2 health centers, and 2 health posts. Mojo City is strategically located along the Ethio-Djibouti railway and was identified by the Ministry as one of the high NCD burden areas in the country.

Study Design

This study used a concurrently embedded a mixed methods approach that combines both quantitative and qualitative data collection strategies at 2 sentinel sites in Mojo City, such as Mojo and Haachaltu Gudina Health Centers, and their 2 administrative institutions, such as the Mojo City health administration office and Oromia regional health Bearau. The evaluation included the performances of these sites from September 1, 2021 to July 30, 2022.

Sampling Procedures

We purposively included the only available 2 sentinel sites in Mojo City, that is, Mojo and Haachaltu Gudina Health Centers. Based on the phenomenon of interest, recommendations from qualitative studies, and practical considerations such as data and resource needs, 14 key informants were purposively selected based on predefined criteria, such as training, experience, education, and position they held.

We included and reviewed the central triage register of all adults aged 30 years and older, screened for hypertension to identify patients with initially elevated BP. We included all actively followed patients with hypertension for BP control assessment. Also, we randomly took a convenient sample of 20 patients' records to assess data completeness and the agreement between cohort register and surveillance reporting platforms. For sensitivity calculation, we compared the proportions of hypertensive cases captured by the surveillance system with those revealed by the survey findings. Finally, we took all individuals having initially raised BP and later confirmed to truly have hypertension to determine positive predictive value (PPV).

Data Collection Procedures

We followed the Centers for Disease Control and Prevention's updated guidelines for evaluating public health surveillance systems to assess the usefulness and 9 attributes of the surveillance system [21]. In addition, we adapted the standards for communicable disease surveillance and response systems evaluation to assess the core and supportive surveillance functions such as case detection, investigation and confirmation, communication and reporting, routine data analysis and interpretation, the availability of surveillance platforms, training, and supportive supervision and feedback system [22].

Data were collected using a semistructured questionnaire through face-to-face interviews, observations, and record reviews. Simplicity, flexibility, stability, acceptability, and usefulness were assessed via key informant interviews. A semistructured interview guide was developed in English and translated into the local language, Oromifa. The interview guide was designed in a way that allowed free responses as well as choices from predefined categories, for example, yes or no response-option questions. After the informant consent was obtained from each key informant, the lead author (ADG) conducted 14 face-to-face interviews. Interviews were

XSL•FO
RenderX

audio-recorded, notes were taken, and each interview lasted 15 to 30 minutes.

We assessed data analysis and interpretation practice, availability of surveillance guidelines and standards, supervision and feedback, and timeliness via observations. Using medical record reviews, hypertension screening and case detection, BP control, investigation and confirmation, BMI screening, data quality, PPV, and representativeness were assessed. Both medical record reviews and observations were conducted by trained Bachelor of Science (BSc) degree nurses.

Operational Definitions

Usefulness

Usefulness refers to the relevance of the surveillance system to detect diseases to injuries, estimate its magnitude, and assess the effects of prevention and control efforts [21]. The system was considered "useful" if it provided accurate and timely data on the magnitude of hypertension in the community, and supported decision-making for health interventions.

Simplicity

Simplicity refers to the ease of the surveillance structures and the ease of operation. A "simple" system was defined as one that required minimal resources, training, time to operate, and avoids complexity about in its information flow.

Acceptability

Acceptability refers to the willingness of surveillance personnel and organizations to participate in the surveillance system. The system was considered "acceptable" if the surveillance staff was willing to implement the system. This was assessed by the status of their active participation in case detection, adherence to case definition, completeness, and timeliness of reporting.

Timeliness

Timeliness reflects the speed between steps in a public health surveillance system. Timeliness was calculated by assessing how many of the expected reports have been submitted within the system-prescribed time frame. The system was considered "timely" if all expected reports have been submitted within a system-defined time frame.

Data Quality

Data quality refers to the completeness and validity of the recorded data in the public health surveillance system. It was assessed in terms of data completeness, expressed as the proportion of reports submitted to the next level of health authority, the completeness of surveillance platforms, measured by the proportion of surveillance forms fully recorded, and the validity of the data recorded.

Flexibility

Flexibility refers to the ability of the surveillance system to adapt to the changing information needs or operating conditions. A system was considered "flexible" if it could easily accommodate changing needs, such as addition of new data variables, modification of data collection process, or adjustment of reporting mechanisms.

Representativeness

Representativeness refers to the ability of surveillance system to describe the occurrence of health-related events by the person, place, and time characteristics. It was assessed by determining on whether the surveillance system provided an accurate description of hypertension in the target population. The system was considered "representative" if it included a balanced representation of the surveillance population.

Stability

Stability refers to the ability to collect, manage, and provide data properly without failure, and the ability to be operational when needed. The system was considered "stable" if it operated without significant downtime or data loss during its restructuring or changing needs.

Sensitivity

Sensitivity refers to the ability of the surveillance system to accurately identify disease or condition under surveillance. The system was considered "sensitive" if it could capture hypertensive cases that actually have a disease. This was assessed by comparing the proportion of patients with hypertension captured by the surveillance system and those identified by the baseline survey findings.

PPV

PPV refers to the proportion of reported cases that actually have the health-related event under surveillance [21]. PPV was assessed by calculating the proportion of surveillance-reported cases that were confirmed as true cases of hypertension. The system was considered "positively predictive" if could truly identify hypertensive cases.

Data Quality Control

Before the actual data collection, we discussed with key informants to ensure that the evaluation was able to address important questions and gather credible evidence. During each visit of data collection, we briefed participants on the purpose and objective of the evaluation, underscoring that it is not an individual performance evaluation. The entire data collection process was closely monitored by research supervisors and principal investigators.

Data Analysis

The quantitative data were coded, cleaned, and analyzed via SPSS software version 25.0 (IBM Corp) while the qualitative data were analyzed via the following methods. First, we transcribed the audio-recorded data and carefully read the text to obtain a general sense of the content. Second, we identified condensed meaning units and developed initial code books. Code books were designed in a way that they are flexible and devised when novel themes emerged. After the open coding, we created subcategories that merged codes with similar meanings, and subcategories were then combined to establish categories. The data were subsequently analyzed manually, narrated, and summarized via thematic analysis.

Ethical Considerations

The study was approved by the Afar Public Health Institute (approval APH/25/21). We also obtained a collaboration letter

XSL•FO
RenderX

from Ethiopia's Ministry of Health, Oromia's Public Health Institute, and Mojo City's Administration Health Office. Before commencement of interview, informed consent was obtained from the participant to ensure their understanding and agreement to the recording process. After each participant understood the purpose, potential risks and benefits of the study, they were orally consented for participation in the study. The participation process was entirely voluntary and no monetary or material compensation was provided to the participants. Such consent and compensation processes were part of the study protocol approved by the ethical approval committee. The data and all collected information were delinked, deidentified, and anonymyzed for confidentiality. They were also kept in a secure file cabinet.

Results

Study Characteristics

The study invited 14 key informants for interviews, and all were willing to provide an interview. Of these key informants, 4 (29% held master's degrees in public health, 4 (29%) were public health officers, and 6 (42%) were BSc degree nurses, of whom 4 (29%) were surveillance officers, and 2 (14%) were facility heads (Table 1).

Table . Key informant characteristics. Evaluation of a sentinel hypertension surveillance system in Mojo city, Eastern Shewa, Oromia, Ethiopia, 2022.

Key informant characteristics	Frequency, n (%)
Title or degree	
Master of Public Health	4 (29)
Public Health Officer	4 (29)
Bachelors of Science degree nurses	6 (42)
Location	
Mojo Health Center	5 (36)
Haachaltu Gudina Health Center	5 (36)
Mojo City Health Office	2 (14)
Oromia Regional Health Bureau	2 (14)
Work position	
Hypertension screening unit	2 (14)
Hypertension focal persons	4 (29)
Surveillance officers	4 (29)
Data Management	2 (14)
Facility heads	2 (14)
Hypertension training	
Trained for hypertension	6 (43)
Untrained for hypertension	8 (57)
Work experience	
Junior level (<2 y)	2 (14)
Intermediate level (2-<5 y)	4 (29)
Senior level (≥5 y)	8 (57)

Description and Operation of the Surveillance System

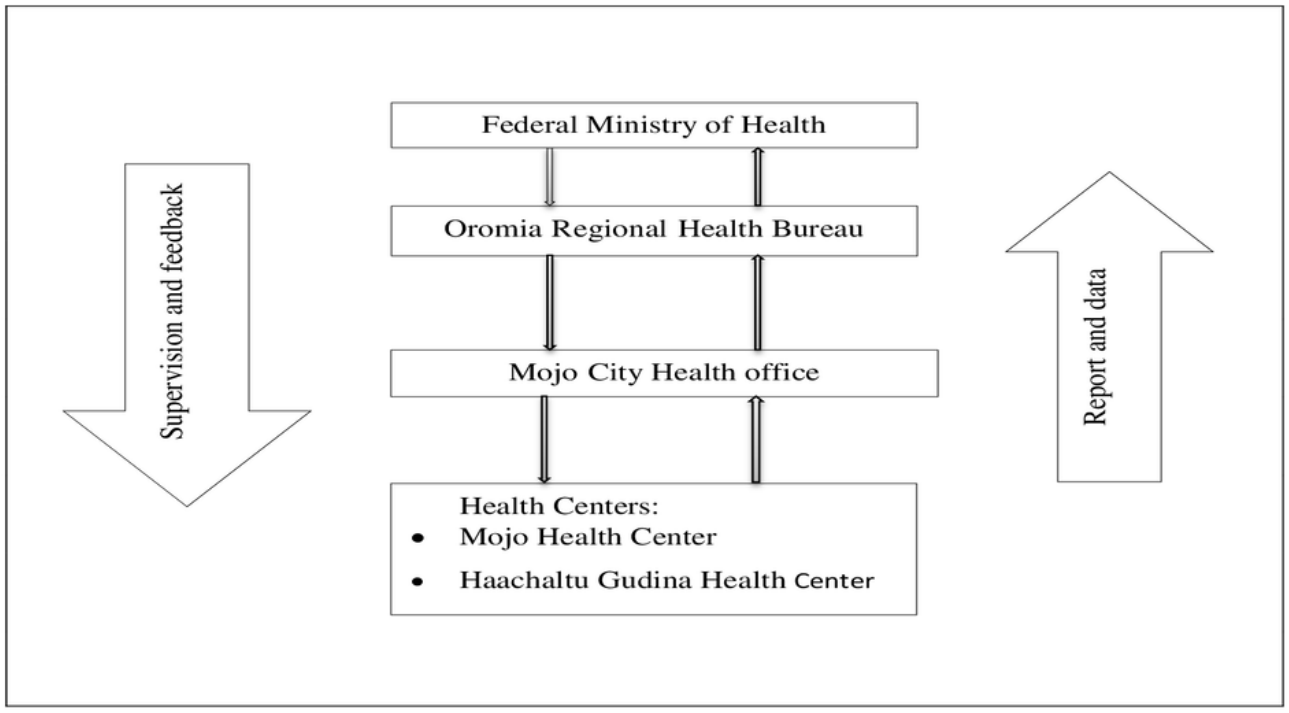
The hypertension surveillance system in the Mojo City is a multitiered framework involving various health care providers and administrative personnel. Public health officers and BSc nurses are responsible for identifying and diagnosing hypertension, and initiating antihypertensive treatment. Health information technicians collect and report data, while the administrative staffs oversee data aggregation, analysis, and interpretation and use for action.

Based on the national clinical guidelines, surveillance system in Mojo City defines hypertension with a systolic blood pressure (SBP) of ≥140 mmHg or a diastolic blood pressure (DBP) of ≥90 mmHg in 2 consecutive visits. BP control is defined as achieving an SBP of <140 mmHg and a DBP of <90 mmHg in 2 consecutive visits, with the specific targets of <130/90 mmHg in patients having comorbidities such as diabetes, chronic kidney disease, and cardiovascular diseases.

The surveillance system uses both electronic and paper-based reporting platforms to facilitate efficient data flow between health care providers and surveillance officers. The information flow follows a bottom-up approach, with data transmitted from health centers to the city health office and then to the regional health bureau (Figure 1).

XSL•FO
RenderX

Figure 1. The information flow and reporting scheme for a sentinel hypertension surveillance system in Mojo city, Eastern Shewa, Oromia, Ethiopia, 2022.



Case Detection

Between September 1, 2021, and July 30, 2022, a total of 3760 adults aged 30 years and older were screened for hypertension. Among those screened, 553 (14.7%) individuals were presented with raised BP, and 512 (13.6%) individuals were confirmed to have hypertension. Of the patients confirmed with hypertension, 76 (14.8%) patients were lost, while 436 (85.2%) were receiving antihypertensive treatment. Only 214 (49%) patients who were on antihypertensive follow-up had achieved controlled BP.

Investigation and Confirmation

A systematic approach was employed to confirm individuals with initially raised BP. Nurses and health officers oversee BP screening and BP confirmation. All patients with raised BP were reassessed for confirmation according to the national clinical guidelines. However, the initiation of antihypertensive therapy lacked essential investigations. Specifically, only 9.7% (50/512) of patients received a baseline renal panel, despite it being a crucial step outlined in the national clinical guidelines and patients may be at higher risk of undiagnosed kidney damage, which could lead to worse outcomes, including kidney failure and cardiovascular complications. The test is essential for evaluating kidney function, detecting early signs of kidney damage, assessing potential comorbidities related to hypertension, such as chronic kidney disease, tailoring treatment plans, and preventing the progression of hypertension-related renal complications. Furthermore, BMI screening, another essential component of hypertension management was not systematically conducted for all patients, with only 25.8% (132/512) of patients with hypertension undergoing this assessment.

Communication and Reporting System

The surveillance system uses Health Management Information System (HMIS) framework of reporting diseases in Ethiopia. Health centers submit reports to Mojo city administration Health Office every 21st-26th of each month. Similarly, Mojo city's administration Health Office subsequently compiled and sent to Oromia Regional Health Bureau within the same time frame. The Oromia Regional Health Bureau further transmits reports to the Federal Ministry of Health within the same reporting schedule. This type of surveillance system was fully implemented throughout the year and maintained through both electronic and paper-based reporting system (Figure 1).

Data Analysis, Interpretation, Action, and Response

Health facilities did not conduct routine data analysis, resulting in hypertension data that lacked specific information regarding time, place, and person characteristics. In addition, there were no established action thresholds for the interpretation of data or for formulating appropriate responses.

Availability of Surveillance Guidelines, Standards, and Other Resources

The surveillance system in the Mojo City ensured the continuous supply of essential surveillance forms, standards, and guidelines. Both health facilities had sufficient back up store of tally sheets, intake forms, follow-up forms, referral forms, and cohort registers. However, health facilities had the shortage of appointment forms and tablets for SIMPLE App (Resolve To Save Lives), a smart app designed for tracking patient follow-up.

Supervision, Feedback, and Training

Neither Ethiopia's Federal Ministry of Health nor the Oromia Regional Health Bureau provided written feedback regarding the performance of health facilities. Only 6 out of 14 (43%) of

XSL•FO
RenderX

key informants received hypertension training. In addition, the heads of health facilities expressed concerns about inadequate training and backup staff to manage staff turnover and annual leave.

Usefulness and Surveillance System Attributes

Usefulness

All 14 key informants (100%) indicated that the surveillance system was valuable, as its record-keeping capabilities documented data that highlighted the magnitude of hypertension in the area and helped them to integrate hypertension service at the community-level involving community health extension workers. One participant noted:

> *I found the existing system immensely important because it helped us understand the extent of the hypertension problem in the area, evaluate our efforts in prevention and control, and improve chart keeping. I believe the lessons learned will be instrumental for further program expansion. [Participant 5, Head of Health Care Site X]*

Acceptability

The active participation of surveillance staff in case detection, adherence to case definitions, and report completeness and timeliness demonstrated their full commitment to implementing the surveillance system. One participant expressed:

> *I am very happy being the hypertension focal person at the health facility and part of this sentinel surveillance system. I am fully committed to hypertension screening and case detection, utilizing case definitions, and providing timely and complete surveillance reports. [Participant 6, Hypertension Focal Person from Facility Y]*

Simplicity

All 14 (100%) of key informants agreed that the current case definition was clear, easy to understand, and apply, and that reporting platforms are user-friendly, requiring average completion time of 10 - 15 minutes. They noted that the route of information flow was clear and well-delineated. They also reported that the SIMPLE App is intuitive and required minimal or no training for data entry and updates. One participant stated:

> *... Without any reservation, in my view, the hypertension case definition is easy for me to understand and apply, and I am able to complete surveillance forms within a matter of 10 minutes. I am also clear with for whom to report, reporting period, and formats used for reporting. [Participant 9, surveillance officer from health care site Y]*

Flexibility

In total, 10 out of 14 key informants (71%) thought that the current reporting form could be easily adapted to new health events, and that case definition of hypertension was flexible enough to accommodate changes. The system has previously experienced significant changes in case definition and its treatment targets, and that they did not face much difficulty in both aspects. This adjustment shifted the focus from individuals aged 18 years and older to those aged 30 years and older for BP screening, as well as updated BP control treatment targets of BP below 130/90 mmHg for patients with hypertension having diabetes, chronic kidney disease, and cardiovascular conditions. One participant stated:

> *I think we can modify the current case definition slightly to allow for other events. Moreover, I believe the current system can operate with varying funding sources, as it was once well established. [Participant 3, Nurse at the Hypertension Screening Unit from Health Care Site X]*

Stability

Out of 14, 7 participants (50%) of the surveillance staff noted that restructuring during the down times of the computer system breakage did not affect the operation of surveillance system and that surveillance system was able to continue to collect, manage, and provide data properly without failure. However, the other half-expressed concerns regarding the adequacy of trained backup staff and the program's dependency on a partner, citing insufficient input from local and regional governments. One participant noted:

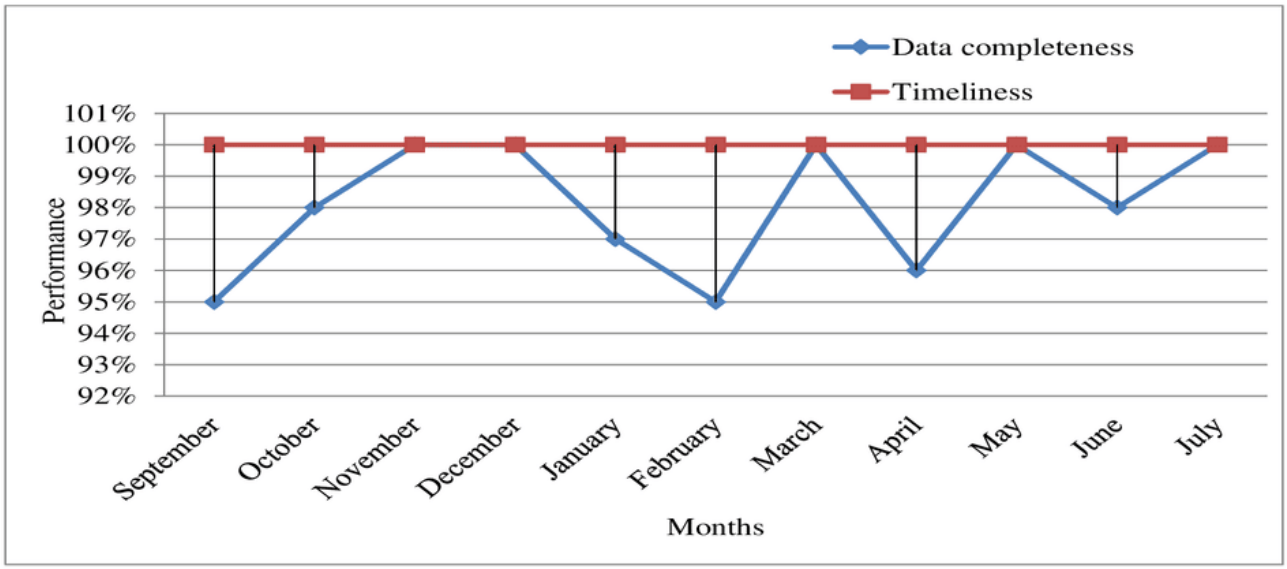
> *There is a common practice in Ethiopia: programs supported by partners and external sources tend to be more effective, but the challenge arises when they are left to operate independently. Our system is similarly dependent on the partner; how could it continue if the partner withdraws support? I have serious concerns about its sustainability. [Participant 2, Surveillance Officer from Administrative Institution B]*

Timeliness

For all 11 months included in the evaluation, reports between each reporting units in the surveillance chain, specifically from health facilities to the Mojo City Administration Health Office, from the Mojo City Administration Health Office to the Oromia Regional Health Bureau, and from the Oromia Regional Health Bureau to the Federal Ministry of Health were submitted according to HMIS framework for reporting diseases in Ethiopia between the 21st and 26th of each month (Figure 2).

XSL•FO
RenderX

Figure 2. The report timeliness and completeness in a hypertension sentinel surveillance system in Mojo City, East Shewa Zone, Oromia region, Ethiopia, 2022.



Data Quality

Of the 20 patient medical records reviewed, 17 (85%) patient charts and follow-up forms were complete, neat, and filled in an easily understandable manner. The agreement between patient charts and cohort register was observed for all these patients. The completeness of the reports over the 11 months of review was 98% (Figure 2).

Representativeness

By screening 3760 adults aged 30 years and older from a corresponding target population of 17,000 inhabitants, the surveillance system reached 22% of the target population. The system identified 512 patients with hypertension, of whom 352 (69%) were reported from the Mojo Health Center, and 270 (52.9%) were females.

PPV and Sensitivity

The BP screening and confirmation practices were taken to calculate PPV. Of the 553 patients with initially raised BP levels, that is, SBP ≥140 mmHg or DBP ≥90 mmHg, 512 (92.6%) individuals truly had hypertension after reassessment (PPV). Similarly, to assess sensitivity, we used a 30% prevalence of hypertension from a baseline facility-based survey conducted in 2020 using a representative sample as a benchmark and compared the current proportion of patients with hypertension identified by the surveillance system (512/3760, 13.6%). Thus, the sensitivity of the surveillance system was 45.3% (Table 2).

Table . A performance comparison table showing system attributes in a hypertension surveillance system in Mojo City, East Shewa Zone, Oromia region, Ethiopia, 2022.

System attributes	Haachaltu Gudina health center	Mojo health center	Mojo city health office	Conclusion
Usefulness	Useful	Useful	Useful	Useful
Acceptability	Acceptable	Acceptable	Acceptable	Acceptable
Simplicity	Simple	Simple	Simple	Simple
Flexibility	Flexible	Flexible	Flexible	Flexible
Stability	Unstable	Unstable	Unstable	Unstable
Representativeness	(1650/7534, 21.9%)	(2110/9466, 22.3%)	(3760/17,000, 22%)	Unrepresentative
Data completeness	(505/520, 97.1%)	(552/559, 98.7%)	(1057/1079, 98%)	High
Timeliness	(11/11, 100%)	(11/11,100%)	(11/11,100%)	High
Sensitivity	43.5%	47.1%	45.3%	Low
Positive predictive value	91.8%	93.4%	92.6%	High

Discussion

Principal Findings

We followed the Centers for Disease Control and Prevention's updated guidelines for evaluating public health surveillance systems to assess on whether the existing surveillance system in Mojo City operated effectively and efficiently. The usefulness and 9 system attributes were streamlined in this evaluation. Accordingly, the surveillance system in Mojo city was useful, simple to operate, flexible to changing needs, acceptable by

XSL•FO
RenderX

surveillance staff, has higher PPV but less sensitive, unrepresentative, and unstable.

While the surveillance system in Mojo City was fully operational throughout the year, it lacked regular supportive supervision and feedback mechanisms. There is need to institutionalize periodic evaluations, enhance the system's functionality, and foster a continuous improvement. Public health surveillance system should be periodically evaluated to provide scientifically sound and epidemiologically interpreted feedbacks for performance and further for its improvement [22,23].

All surveillance staff expressed willingness to implement the system and noted that it was both useful and simple. Half perceived the surveillance system stable, while 71% (10/14) thought that it was flexible. Similarly, the performance of public health surveillance system in Dangila, Ethiopia, was found flexible and useful [24]. Furthermore, a hypertension surveillance system evaluation conducted in Indonesia at 2 health centers and 1 administrative office reported an overall low stability, simplicity, acceptability, and data quality, but the researchers were unable to assess flexibility [25]. However, unlike this study, researchers relied on data collection, storage, and reporting to assess stability, rather than using key informant interviews [25].

Given that 3760 adults aged ≥30 years were screened for hypertension out of a corresponding population of 17,000 inhabitants, the surveillance system in a Mojo City demonstrated low representativeness, having reached only 22% of the population. However, in a hypertension surveillance evaluation conducted in Indonesia, they were unable to assess representativeness of surveillance system [25]. The low representativeness in the present surveillance system may be attributed to low community awareness and routine BP screening, and regular medical checkup tradition, which could leave significant numbers of undiagnosed patients in the community. In this system, health centers collect data and report it to the Mojo City Health Office and the Oromia Regional Health Bureau. However, health facilities did not conduct routine data analysis or interpretation, nor did they use the information to inform action. Data collection necessitates analysis because the goal of surveillance is to use the information for action, reduce morbidity and mortality, and to improve overall health [22,23]. If the collected data are not followed by regular analysis and the application of information for action, it results in the mere consumption of resources [26].

Public health surveillance systems should operate in a manner that facilitates effective communication and dissemination of health and health-related data, enabling decision-makers at all levels to readily understand the implications of the information. Timeliness of reporting is a crucial factor for the effective communication and dissemination of health and health-related data. The report timeliness in the current surveillance system was 100%, which aligns with the Ethiopian national target of 95% [27] and the WHO target of 80% [28]. In addition, data completeness was 98%, consistent with the national target of over 95% for report completeness [27] and the WHO target of 80% [28].

The sensitivity and PPV of a surveillance system depend on the design of the case definition and its clarity among stakeholders [29]. The surveillance system in Mojo City exhibited a sensitivity of 45.3%, capturing only 45.3% of hypertension cases and more than half (54.7%) of the individuals with hypertension may be missed by the system. It is less sensitive than the hypertension surveillance system in Bogor City, Indonesia, which has a sensitivity of over 80% [30]. The low sensitivity may be attributed to low health care service utilization and suboptimal BP screening practice, as only 22% of the surveillance population was screened for BP. In a hypertension surveillance system evaluation at the Jombang District Health Office, Indonesia, they were unable to calculate sensitivity due to the lack of baseline survey data for comparison. However, the PPV of our surveillance system was 92.6%, indicating that the system correctly identifies 92.6% of true cases of hypertension. It is more predictive when compared with other hypertension surveillance systems in Dangila, Ethiopia which has PPV of 11% [24] and Jombang District Health Office of the other city in Indonesia which has PPV of 59.4% [25].

Limitations of the Study

The limitations of this evaluation are as follows. First, the results may have been subject to potential social desirability bias, as surveillance stakeholders may have provided favorable responses especially qualitative attributes such as acceptability and simplicity. The surveillance stakeholders may have been inclined to provide more favorable responses to questions about these attributes, potentially overestimating the true level of satisfaction or ease with the surveillance system. To mitigate this, we briefed them on the purpose and objectives of the evaluation at each visit before commencing data collection. Second, reliance on a convenience sample of 20 patients may not have provided a representative assessment of data validity. Third, although this evaluation followed the Centers for Disease Control and Prevention's updated guidelines for evaluating public health surveillance systems as applied in previous studies conducted in different settings to gather credible evidence on the performance of the surveillance system, it is primarily relied on responses from key informant interviews rather than objective quantitative data collection, storage, and reporting, unlike other studies. Finally, the availability of the limited data in the field led to insufficient understanding of surveillance systems in similar contexts and a few of interpretations were made compared with surveillance systems in other country which may have structural and contextual difference with our surveillance system.

Conclusion

The sentinel hypertension surveillance system in Mojo demonstrated effective implementation throughout the year. The surveillance staff felt that it was helpful in determining the magnitude of hypertension in the area as well as for improving documentation and patient record keeping practices. It was found to be simple, flexible, acceptable, and have a higher PPV. The data quality and reporting timeliness were consistent with national and WHO surveillance standards. However, it was less sensitive, less representative, and unstable.

XSL•FO
RenderX

The findings underscore the need for institutionalization of periodic evaluations and supportive supervision mechanisms to assess the system's performance, provide feedback, and identify areas for improvement. The system should implement routine data analysis and interpretation mechanisms based on epidemiological principles to interpret the collected data, identify trends, and target tailored public health interventions. Continuous staff training on data collection, surveillance protocols, and the operation of the surveillance system should be designed to enhance stability. Furthermore, designing of government-driven sustainability strategies, such as establishment of dedicated public health financing system and consistent resource allocation and capacity building, and cross-sectoral collaborations with local governments, could reduce reliance on external donors and enhance the sustainability of the surveillance system. Efforts should also focus on community awareness creation activities to improve sensitivity and representativeness of the surveillance system. Finally, if resources permit, scaling the system into a nationwide program could significantly enhance hypertension surveillance across Ethiopia.

Acknowledgments

The authors would like to thank Ethiopian Field Epidemiology Training Program, Oromia Public Health Institute, Mojo City Administration Health Office, Mojo and Haachaltu Gudina health centers, health workers, and data collectors.

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee support for the publication of this article.

Authors' Contributions

ADG conceived the research idea, designed and conducted the study, acquisition of data, analyzed the data, and drafted the article. DZW provided substantial contributions to the design of the study and drafted the article. YAF conducted an extensive manuscript review and structuring. AYE provided substantial contributions to conception and design, or analysis and interpretation of data; and drafting the article or revising it critically for important intellectual content. All the authors have read and approved the final manuscript contributed to conception and design.

Conflicts of Interest

None declared.

References

1. Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. J Am Coll Cardiol 2019 Nov 19;74(20):2529-2532. [doi: 10.1016/j.jacc.2019.10.009] [Medline: 31727292]
2. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. Nat Rev Nephrol 2020 Apr;16(4):223-237. [doi: 10.1038/s41581-019-0244-2] [Medline: 32024986]
3. Hypertension key fact sheet. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/hypertension [accessed 2021-08-01]
4. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. The Lancet 2005 Jan;365(9455):217-223. [doi: 10.1016/S0140-6736(05)17741-1]
5. Zhou B, Carrillo-Larco RM, Danaei G, et al. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. The Lancet 2021 Sep;398(10304):957-980. [doi: 10.1016/S0140-6736(21)01330-1]
6. Molla M. Systematic reviews of prevalence and associated factors of hypertension in Ethiopia: finding the evidence. SJPH 2015;3(4):514. [doi: 10.11648/j.sjph.20150304.19]
7. Shiferaw F, Letebo M, Misganaw A, Feleke Y, Gelibo T, Getachew T, et al. NCDs in Ethiopia: Disease burden, gaps in health care delivery and strategic directions. Ethiopian Journal of Health Development 2018;32(3).
8. Misganaw A, Mariam DH, Ali A, Araya T. Epidemiology of major non-communicable diseases in Ethiopia: a systematic review. J Health Popul Nutr 2014 Mar;32(1):1-13. [Medline: 24847587]
9. Jung M, Jembere GB, Park YS, et al. The triple burden of communicable and non-communicable diseases and injuries on sex differences in life expectancy in Ethiopia. Int J Equity Health 2021 Aug 3;20(1):180. [doi: 10.1186/s12939-021-01516-0] [Medline: 34344371]
10. Federal Democratic Republic of Ethiopia, Ministry of Health. National strategic action plan (NSAP) for prevention and control of noncommunicable diseases in ethiopia, 2014-2016. URL: https://www.iccp-portal.org/system/files/plans/ETH_B3_National%20Strategic%20Action%20Plan%20%28NSAP%29%20for%20Prevention%20and%20Control%20of%20Non-Communicable%20Diseases%20%20Final.pdf [accessed 2025-05-29]
11. World Health Organization. Non-Communicable Diseases Country Profiles 2018. URL: https://www.who.int/publications/i/item/ncd-country-profiles-2018 [accessed 2022-08-23]

XSL•FO
RenderX

12. Legese N, Tadiwos Y. Epidemiology of hypertension in Ethiopia: a systematic review. Integr Blood Press Control 2020;13:135-143. [doi: 10.2147/IBPC.S276089] [Medline: 33116810]
13. Kebede B, Chelkeba L, Dessie B. Rate of blood pressure control and its determinants among adult hypertensive patients at Jimma University Medical Center, Ethiopia: Prospective cohort study. SAGE Open Med 2021;9:20503121211006000. [doi: 10.1177/20503121211006000] [Medline: 34659760]
14. Horsa BA, Tadesse Y, Engidawork E. Assessment of hypertension control and factors associated with the control among hypertensive patients attending at Zewditu Memorial Hospital: a cross sectional study. BMC Res Notes 2019 Mar 18;12(1):152. [doi: 10.1186/s13104-019-4173-8] [Medline: 30885272]
15. Gebremichael GB, Berhe KK, Zemichael TM. Uncontrolled hypertension and associated factors among adult hypertensive patients in Ayder comprehensive specialized hospital, Tigray, Ethiopia, 2018. BMC Cardiovasc Disord 2019 May 22;19(1):121. [doi: 10.1186/s12872-019-1091-6] [Medline: 31117945]
16. Animut Y, Assefa AT, Lemma DG. Blood pressure control status and associated factors among adult hypertensive patients on outpatient follow-up at University of Gondar Referral Hospital, northwest Ethiopia: a retrospective follow-up study. Integr Blood Press Control 2018;11:37-46. [doi: 10.2147/IBPC.S150628] [Medline: 29720880]
17. WHO. Noncommunicable Disease Surveillance, Monitoring and Reporting Surveillance of noncommunicable diseases. URL: https://www.who.int/teams/noncommunicable-diseases/surveillance [accessed 2021-07-20]
18. Ethiopian Health and Nutrition Research Institute. Public Health Emergency Management Centre Public Health Emergency Management guidelines for Ethiopia, 2012.: Addis Ababa; 2012. URL: https://www.medbox.org/document/public-health-emergency-management-guidelines-for-ethiopia-2012 [accessed 2025-05-29]
19. Ethiopian Public Health Institute Public Health Emergency Management Center, Federal Democratic Republic of Ethiopia. Public Health Emergency Management Guideline for Ethiopia, 2nd edition 2022. URL: https://eweb.ephi.gov.et/wp-content/uploads/2024/08/PHEM_Guideline_Second_Edition_2023.pdf [accessed 2025-05-29]
20. WHO regional office for Africa. Ethiopia sets to improve hypertension prevention and control at primary health care level. 2019 Aug 16. URL: https://www.afro.who.int/news/ethiopia-sets-improve-hypertension-prevention-and-control-primary-health-care-level [accessed 2021-08-25]
21. Centers for Disease Control and Prevention (CDC). Updated guidelines for evaluating public health surveillance systems recommendations from the guidelines working group. 2001 Jul 27 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm [accessed 2021-07-01]
22. WHO. Communicable disease surveillance and response systems. guide to monitoring and evaluating. : World Health Organizaton; 2006 URL: https://iris.who.int/bitstream/handle/10665/69331/WHO_CDS_EPR_LYO_2006_2_eng.pdf [accessed 2021-08-01]
23. LANGMUIR AD. The surveillance of communicable diseases of national importance. N Engl J Med 1963 Jan 24;268(4):182-192. [doi: 10.1056/NEJM196301242680405] [Medline: 13928666]
24. Alemu T, Gutema H, Legesse S, Nigussie T, Yenew Y, Gashe K. Evaluation of public health surveillance system performance in Dangila district, Northwest Ethiopia: a concurrent embedded mixed quantitative/qualitative facility-based cross-sectional study. BMC Public Health 2019 Oct 22;19(1):1343. [doi: 10.1186/s12889-019-7724-y] [Medline: 31640662]
25. Lestari D. Evaluation of monitoring hypertension case based on nine surveillance attributes in Jombang District Health Office in 2018. ijph 2021;16(1):89. [doi: 10.20473/ijph.v16i1.2021.89-100]
26. Foege WH, Hogan RC, Newton LH. Surveillance projects for selected diseases. Int J Epidemiol 1976 Mar;5(1):29-37. [doi: 10.1093/ije/5.1.29] [Medline: 944166]
27. The Federal Democratic Republic of Ethiopia, Ministry of Health Health Sector Transformation Plan. Health Sector Transformation Plan 2015/16 - 2019/20 (2008-2012 EFY) August 2015. Food and Agriculture Organization of of the United Nation. URL: https://www.cmpethiopia.org/content/download/2268/9612/file/HSTP-Final-2015-10-19.pdf [accessed 2021-03-21]
28. World Health Organization. Integrated disease surveillance and response technical guidelines: booklet three: sections 4, 5, 6, and 7. : World Health Organisation; 2019 Mar URL: https://apps.who.int/iris/bitstream/handle/10665/312362/WHO-AF-WHE-CPI-03.2019-eng.pdf [accessed 2021-04-05]
29. PennState Eberly College of science. STAT 507 | Epidemiological Research Methods. STAT ONLINE | Department of Statistics. URL: https://online.stat.psu.edu/stat507/lesson/2/2.1 [accessed 2021-07-08]
30. Amalia D, Kusnadi B, Bantas K. The evaluation of hypertension surveillance system based on primary health care in bogor city. Indonesia: The 8th National Scientific Conference on Epidemiology.: Faculty of Public Health, Universitas Indonesia URL: https://www.researchgate.net/publication/337293267_The_Evaluation_of_Hypertension_Surveillance_System_Based_on_Primary_Health_Care_in_Bogor_City_2018 [accessed 2025-05-29]

Abbreviations

BP: blood pressure
BSc: Bachelors of Science

XSL•FO
RenderX

DBP: diastolic blood pressure
HMIS: Health Management Information System
NCD: noncommunicable disease
PPV: positive predictive value
SBP: systolic blood pressure
WHO: World Health Organization

Edited by E Mensah; submitted 20.02.25; peer-reviewed by A Adekola, C Udensi; revised version received 24.03.25; accepted 10.04.25; published 03.06.25.

Please cite as:
Demelash Gashe A, Ayale Ferede Y, Zenebe Weldemichael D, Yesuf Endries A
Evaluation of a Sentinel Hypertension Surveillance System in Mojo, East Shewa Zone, Oromia, Ethiopia: Concurrently Embedded Mixed Design Study
Online J Public Health Inform 2025;17:e72909
URL: https://ojphi.jmir.org/2025/1/e72909
doi:10.2196/72909

© Abiyie Demelash Gashe, Yeshiwas Ayale Ferede, Dawit Zenebe Weldemichael, Aman Yesuf Endries. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 3.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cultivating Disaster Preparedness: Scoping Review of Technology's Contribution to Situational Awareness and Disaster Mindset in Disaster Medicine

Amir Khorram-Manesh[1], MD, PhD; Marius Rohde Johannessen[2], PhD; Laurits Rauer Nielsen[3], MSc; Eric Carlström[4], PhD; Lasse Berntzen[5], PhD; Lene Sandberg[3], MEd; Jarle Løwe Sørensen[2], DBA

[1]Surgery, Clinical Sciences, University of Gothenburg, Universitetsplatsen 1, Gothenburg, Sweden

[2]Department of Business, History and Social Sciences, USN School of Business, University of South-Eastern Norway, Borre, Norway

[3]Management and Administration Emergency and Risk Management, University College Copenhagen, Copenhagen, Denmark

[4]Learning and Leadership for Health Care Professionals, Institute of Health and Care Sciences, University of Gothenburg, Gothenburg, Sweden

[5]Department of Business, Marketing and Law, USN School of Business, University of South-Eastern Norway, Borre, Norway

Corresponding Author:
Amir Khorram-Manesh, MD, PhD
Surgery, Clinical Sciences, University of Gothenburg, Universitetsplatsen 1, Gothenburg, Sweden

Abstract

Background: Disaster medicine education increasingly emphasizes situational awareness and a proactive disaster mindset as crucial competencies for effective response. Situational awareness involves comprehending the disaster environment to make informed decisions under pressure, while a disaster mindset encompasses psychological resilience and effective functioning amid chaos. Integrating technologies into simulation training allows experiential learning that bridges these theoretical concepts with practical application.

Objective: This study aims to investigate the current status of teaching these concepts and the use of technology in fostering situational awareness and a disaster mindset within disaster medicine education by reviewing the existing literature.

Method: This study used a scoping review of scientific studies (2005 - 2025), obtained from PubMed, Scopus, and Web of Science databases, complemented by a Google Scholar search. From December 1, 2024, to the end of January 2025, 3 reviewers searched, compiled, reviewed, and selected eligible studies in English, discussing the use of technology in fostering situational awareness and disaster mindset.

Results: Out of 217 initially identified records, 49 studies met the inclusion criteria after a 2-stage screening and full-text review process. Of these, 42 were peer-reviewed scientific articles and 7 were official documents. Approximately 86% (42/49) of the studies addressed situational awareness, while only 2% (1/49) explicitly focused on the concept of disaster mindset. Most of the included studies highlighted the use of immersive technologies such as virtual and augmented reality, geographic information systems, and artificial intelligence–driven tools to enhance real-time information processing and decision-making in disaster education contexts. By strategically incorporating these advanced tools into educational frameworks, the divide between theoretical knowledge and practical application can effectively be bridged, fostering essential experiential learning and developing robust psychological readiness for future challenges.

Conclusions: Simulation training enhances situational awareness and disaster mindset, bridging the gap between theory and practice through experiential learning. The findings from this review highlight current pedagogical approaches and technological applications, identifying gaps and future directions for enhancing disaster medicine education.

(Online J Public Health Inform 2025;17:e75404) doi:10.2196/75404

KEYWORDS

disaster education; disaster medicine; disaster mindset; simulation exercises; situational awareness

Introduction

The 21st century has experienced a significant transformation in disaster medicine, disaster management, and public health, primarily driven by the rapid advancement of technology. Since the early 2000s, technology has played a pivotal role in shifting these fields from reactive to proactive approaches and from localized to globally interconnected systems [1-3].

Initially, the advent of the internet and mobile communication facilitated faster dissemination of information and coordination during crises [4]. Real-time updates, emergency alerts, and remote communication became essential tools for first

XSL•FO
RenderX

responders and public health officials. Geographic information systems (GIS) emerged as a powerful instrument for mapping disaster zones, tracking disease outbreaks, and optimizing resource allocation, providing unprecedented spatial awareness [5]. As technology progressed, its applications expanded. Remote sensing and satellite imagery enabled the early detection of natural hazard-induced disasters, facilitating preemptive evacuations and resource mobilization [6]. The rise of social media platforms introduced new channels for public communication, citizen reporting, and real-time information gathering, although they also presented challenges related to misinformation [7].

The proliferation of mobile health apps and wearable devices has revolutionized public health monitoring and disease surveillance [8]. Real-time data collection, remote patient monitoring, and telehealth services became integral to managing chronic conditions and responding to outbreaks, particularly in resource-limited settings. Furthermore, the development of simulation and virtual reality (VR) technologies transformed training and education in disaster medicine. Immersive simulations allowed health care providers to practice complex procedures and decision-making in realistic, high-stress environments, thereby enhancing preparedness and competency [9].

Artificial intelligence (AI) and machine learning algorithms have emerged as powerful tools for analyzing vast datasets, predicting disease outbreaks, and optimizing disaster response strategies. AI-powered chatbots and online assistants provide instant access to information and support, while robotic systems assist in search and rescue operations [10]. In recent years, the COVID-19 pandemic accelerated the adoption of digital health solutions, underscoring the critical role of technology in pandemic preparedness and response. From contact tracing apps to vaccine distribution platforms, technology has been instrumental in managing the global health crisis [11,12].

While challenges persist, including issues related to data privacy, digital equity, and the ethical use of AI, the transformative role of technology in disaster medicine and public health management is undeniable [13]. By continuing to innovate and integrate new technologies, we can build more resilient communities and enhance global health security.

The current inconsistencies in disaster medicine education stem from the diverse risks, causes, needs, resources, experiences, and expertise inherent to different regions, highlighting a critical need for greater collaboration and a more standardized approach to training. Previous research underscores this requirement, pointing to significant variations in elements of the management systems, including command, control, and communication protocols, alongside notable deficiencies in existing training methodologies [14]. Although disaster medicine education is crucial, it suffers from a lack of global consistency in both content and opportunities, with one major issue being the absence of standardized curricula that adequately address the multifaceted nature of disaster response, including emerging technologies and the complexities of hybrid warfare [14-16]. Insufficient emphasis on practical, hands-on training further hinders effective disaster preparedness, compounded by limited access to advanced simulation and high-fidelity training environments that restrict the ability to provide realistic learning experiences [17]. The high cost of these vital technologies often limits their availability to well-funded institutions, leaving many learners without essential training tools. Moreover, greater interagency collaboration is needed in training programs to ensure all stakeholders, including emergency responders, health care professionals, and government officials, are prepared to coordinate effectively during crises [14-16]. Addressing these shortcomings is essential to improve the effectiveness of disaster medicine education and enhance overall preparedness [14-17].

Globally, the use of simulation exercises in disaster education is recognized. These exercises range from tabletop simulations, which require fewer resources and limited financial support, to live or modular simulations, which demand significantly more resources and funding [18]. In theory, simulation exercises are a potent tool in disaster medicine education, enhancing both learning and practical application [19]. However, they predominantly focus on local and regional group and team activities, while interorganizational, transnational, and global collaboration is still in need of development [19,20]. On the one hand, group dynamics may influence personal beliefs and knowledge, potentially leading to resistance and noncompliance with established guidelines and rules of engagement [21]. On the other hand, routines and technology tend to remain intraorganizational, as well as at local, regional, and national levels [22,23].

These challenges highlight the disconnect between the different management levels, probably stemming from a lack of overall situational awareness (SA) and a disaster mindset (DMS). SA involves the ability to accurately perceive, understand, and respond to elements within a given environment, especially in dynamic and high-stress situations [24]. It encompasses being aware of one's surroundings and predicting potential changes or threats. A DMS refers to the mental preparedness and resilience necessary to effectively manage and respond to disaster scenarios [25]. It includes maintaining composure, making rapid and confident decisions, and adapting to extreme stress and uncertainty.

The importance of SA and a DMS is increasingly recognized in disaster medicine, necessitating the incorporation of both in current programs and educational initiatives, through lectures, simulations, and field exercises to better prepare responders for the complex psychological and operational challenges they will face. Many programs may use simulations to replicate disaster scenarios, allowing participants to practice SA and decision-making under stress in a controlled environment. In addition, effective disaster response requires teamwork and communication across different disciplines. Training often emphasizes these aspects, which are closely linked to SA and a collaborative mindset [26-32]. Nevertheless, to the best of our knowledge, SA and DMS are not adequately integrated into current educational initiatives in disaster medicine.

This study aims to investigate the status of teaching these concepts and the use of technology in fostering SA and a DMS, both individually and collectively, as well as locally, regionally, and transnationally, by reviewing the existing literature.

XSL•FO
RenderX

Methods

Study Method

To comprehensively map the broad field of disaster medicine education, a scoping review was conducted (December 2024 to the end of January 2025), using academic databases (PubMed, Scopus, and Web of Science [WoS]) and Google Scholar in several steps, including search, review, and selection of the literature. To effectively research how technology enhances SA and DMS, a combined search strategy is crucial. Google Scholar excels at identifying a broad spectrum of content, including recent and interdisciplinary work. In contrast, specialized scientific databases offer greater precision through curated, peer-reviewed content, controlled vocabularies, and advanced filters [33]. Using a multidatabase strategy allowed for the capture of a broad research spectrum, providing a robust, unbiased overview of disaster medicine education and ensuring the retrieval of all relevant literature, overcoming the limitations of each when used in isolation. This method also aimed to establish a foundational understanding of the field's current state by identifying core concepts, theories, and evidence, with particular attention to SA, DMS, and technology's impact. The review clarified contested definitions, pinpointed critical research gaps, and evaluated the feasibility of a future systematic review [34]. This preliminary assessment serves to guide future research and development in disaster medicine education.

Research Question

How can current technology help enhance SA and DMS in disaster medicine education?

Searching Keywords

Disaster AND Technology AND Situational awareness AND Disaster mindset were selected as the final search keywords and deemed highly adequate for investigating the role and impact of technology on SA and DMS in disaster medicine. Disaster is the core subject of this investigation and broadly encompasses the events (natural or human-made) that necessitate disaster medicine and sets the context for all other keywords, avoiding the capture of SA and DSM in nondisaster contexts (eg, routine health care). It ensures that the obtained literature directly addresses the unique challenges and demands of crisis situations. Technology covers a vast array of innovations relevant to disaster medicine, including but not limited to communication technologies (satellite phones, mesh networks, and mobile apps for coordination), information systems (GIS, data analytics, and data processing), remote tools (telemedicine, drones, remote sensors, and wearable devices), and training and simulation (VR and augmented reality [AR]), ensuring the search captures how these tools are being applied and studied within the disaster medicine context. There are several synonyms to use for both SA and DMS, such as perception, comprehension, and cognition for SA, and disaster readiness and resilience for DMS. However, listing all synonyms could potentially lead to information overload, while targeting the core keywords could yield more targeted results.

Educational programs were not included in the main search and were specifically searched through official webpages. The reasons were as follows: (1) adding "educational programs" makes the search even more specific, (2) many studies on technology, SA, and DMS in disaster medicine inherently involve education and training, even if they don't explicitly use "educational programs" in their keywords or title, (3) educational programs can be very broad, for example, public education and might not involve technology, and (4) their inclusion in every search will constantly result in studies focusing on training. While training is a key aspect, we aimed to capture other ways technology influences SA and DMS (eg, real-time data feeds, decision support systems used in the field, and psychological impacts of constant connectivity).

Databases

To comprehensively explore the scientific literature, this review used PubMed, Scopus, and WoS databases, each offering distinct advantages. PubMed, crucial for biomedical research, provides access to clinical and health-related studies. Scopus, with its interdisciplinary coverage, facilitated the identification of research trends and influential studies across diverse fields. Finally, WoS, known for its rigorous indexing, ensured the inclusion of high-impact research, including ethical and social perspectives.

Search String

Two similar search strings (mentioned below) were used for each database and Google Scholar. SA and DMS were isolated to curtail the results and obtain eligible studies. Using SA and DMS together decreased the number of hits.

Disaster AND "Technology" AND "Situational Awareness"

Disaster AND "Technology" AND "Disaster Mindset"

Inclusion Criteria

Studies in English discussing the impact of SA and DSM in disasters and public health emergencies between 2005 - and 2025 were included.

Exclusion Criteria

Studies not discussing the use of technology or unrelated to the educational aspects of disaster medicine, and those that did not follow the inclusion criteria or were in other languages, were excluded.

Eligibility of Included Studies

Scientific or nonscientific studies that discussed the keywords in English and followed the inclusion criteria were all eligible for inclusion.

Data Extraction

Data were extracted based on a predesign template by 2 of the authors and were validated by a third one. Data from included sources were charted using calibrated forms developed and piloted by the review team prior to use and consisted of title, date of publication, journal, study setting, and key points. This ensured consistency and accuracy in data collection. Data charting was performed independently during the review process, described above. Where necessary, efforts were made to contact original investigators to clarify or confirm data points, adhering to established systematic review guidelines.

XSL•FO
RenderX

Review Process

The review process consisted of 2 stages: initial screening of titles and abstracts, followed by a second review and detailed examination of selected full-text studies, from which data were extracted for inclusion in the results. Three reviewers (AK-M, JLS, and MR-J), one trauma and disaster medicine specialist, one crisis manager, and one technology specialist examined the compiled studies independently to achieve unanimous consensus on inclusion. Any disagreement was resolved by consulting a fourth reviewer.

Content Analysis

The outcomes were divided into diverse themes using conceptual content analysis, based on the similarities and differences [35]. The process includes defining the research question, coding the text, deciding the level of analysis, and developing rules for consistency. Finally, conclusions were drawn based on the patterns and trends identified.

Results

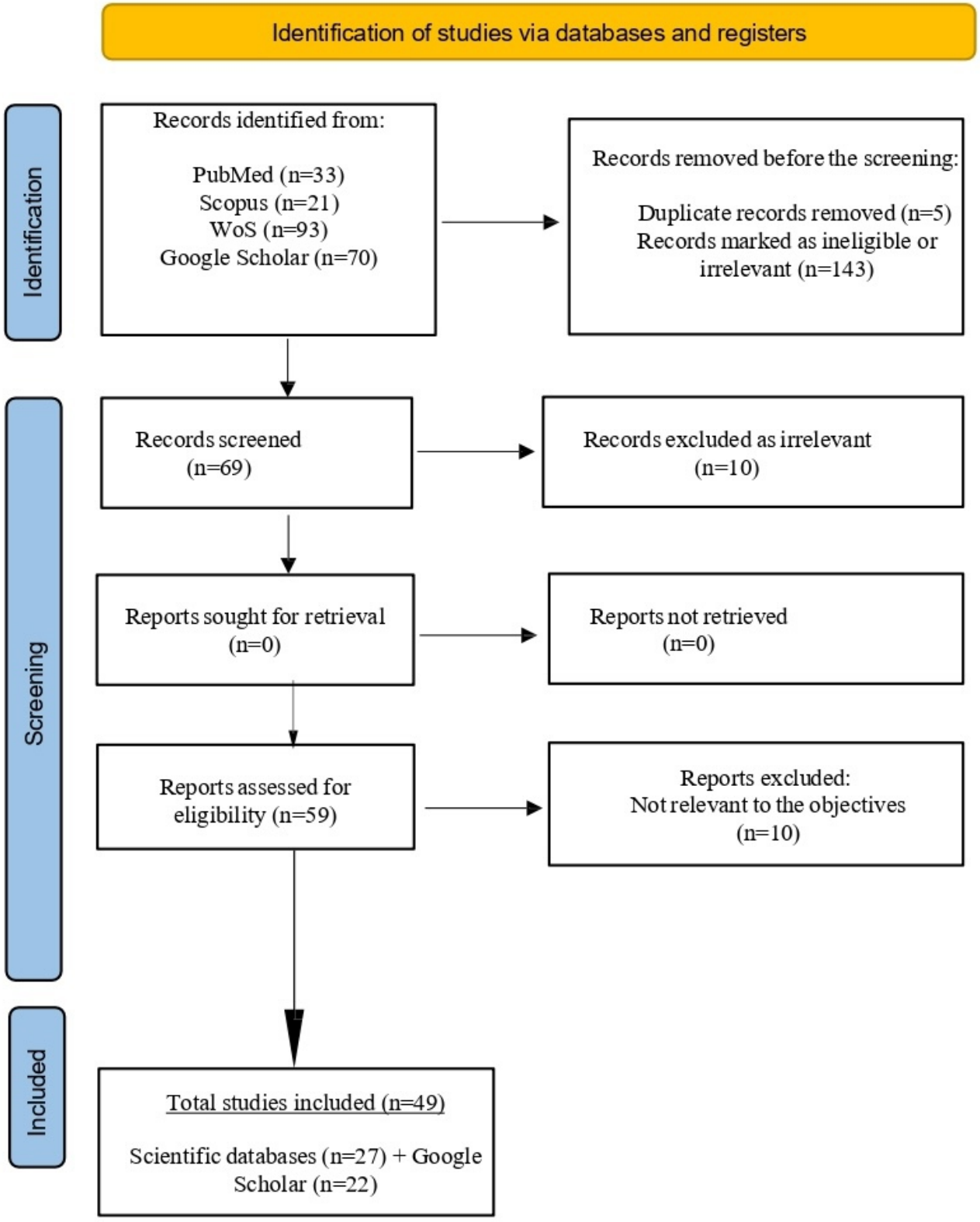
Overview

The search in scientific databases resulted in 147 studies from the following databases: PubMed (n=33), Scopus (n=21), and WoS (n=93). The Google Scholar search resulted in numerous documents. After sorting by relevance, the first 70 hits were added to the review list. Of the total number of 217 documents, in the first step, duplicates (n=5) and irrelevant papers (n=143) were removed by reviewing the titles and abstracts of compiled studies. Sixty-nine studies were considered relevant and included in the second review round by 3 reviewers. Each reviewer studied the included papers separately and selected the final number of studies to be included in this review (Multimedia Appendices 1 and 2 show the strategy and findings of each database and Google Scholar).

Ten irrelevant studies were removed before eligibility assessment. An additional 10 studies were not eligible, and the remaining 49 studies [24,25,36-82] were unanimously selected and included in this study at the final stage (Table 1; Figure 1). The search result consisted of peer-reviewed articles, policy documents, blogs, and guidelines.

XSL•FO
RenderX

Figure 1. The search and literature selection process according to the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flowchart for scoping reviews.



XSL•FO
RenderX

Table . The setting and key points of the included studies, presented after the in-text citation.

Study details and reference number	Study setting	Key points
Mohsin et al (2016) [24]	General disaster response	Focuses on innovative systems for SA[a]
Su et al (2022) [25]	Global COVID-19 pandemic	Emphasizes the importance of a proactive DMS[b]
Naor and Laor (2020) [36]	Puerto Rico, Hurricane Maria recovery	Endsley's model assesses SA in disaster recovery
Kettelhut et al (2017) [37]	Tactical biosurveillance	Visualizations impact biosurveillance SA
Glick and Barbara (2013) [38]	General disaster response	Examines the transition from awareness to decision-making
Jokela et al (2012) [39]	Major incidents	Evaluate RFID[c] technology for improved situational awareness
McCurdy et al (2005) [40]	Medical disaster response	Explores video streaming for enhanced medical response
Kedia et al (2022) [41]	General disaster response	Reviews technologies for SA
Saadou and Chenji (2018) [42]	Disaster response networks	Presents research methods to optimize SA
Yuan et al (2013) [43	Disaster reliefs	Surveys how human sensor networks improve SA
Karavarsamis et al (2022) [44]	Disaster sites, adverse weather	Deep learning improves image clarity for better SA
Jain et al (2021) [45]	Mass-gathering events	Compares UAV[d] effectiveness to standard practices in scene assessment
Jain et al (2018) [46]	Mass casualty incidents	Compares UAV effectiveness to standard practices in hazard identification
Lee et al (2022) [47]	Unmanned aerial vehicles	Explores deep learning for UAV-based human posture recognition
Mogaka et al (2024) [48]	Aerial scene image classification	Created a lightweight deep learning model for aerial image classification
Spandonidis et al (2021) [49]	Air cargo operations	Describes a data-driven system for situational awareness
Yuan et al (2022) [50]	China	Describes the key technologies used in China's emergency response platforms
Tapia et al (2016) [51]	Public universities	Improves 911 texting for large-scale disasters
Gargees et al (2016) [52]	Incident support	Describes the use of visual cloud computing and SDN[e] to support incident response
Chan and Purohit (2020) [53]	Public health systems during disasters	Highlights the challenges of using social media data for public health
Abid et al (2021) [54]	General disaster management	Explores AI's[f] role in improving disaster management
Panah et al (2024) [55]	General disaster zones	Social network analysis enhances disaster health care and resilience
Pekar et al (2020) [56]	General disaster zones, social media	Studies the use of social media for early disaster detection
Zafar et al (2024) [57]	Search and Rescue operations	Uses gesture control and deep learning to enhance unmanned ground vehicle performance in SAR[g]
Lenert et al (2011) [58]	Mass casualty settings	Focuses on wireless EHR[h] systems in field care
Gao et al (2014) [59]	General emergency settings	Focuses on measuring and analyzing information flow
Gandhi et al (2021) [60]	Educational setting	Pandemic simulations improve student preparedness

XSL•FO
RenderX

Study details and reference number	Study setting	Key points
Stroud et al (2010) [61]	Workshop and conceptual setting	Emphasizes the importance of SA in managing medical surge capacity during disasters. Discusses frameworks and challenges in achieving SA
Iserson KV [62]	Conceptual setting	Discusses the integration of various disaster response tools to empower clinical leadership and improve decision-making and coordination
O'Brien et al (2020) [63]	Conceptual setting. Multiagency emergency response	Explores the complexities and identifies challenges of achieving shared SA across different agencies during emergency response
Schwarz et al (2023) [64]	Case study	Investigates the potential for automating SA reporting in disaster management
Zhai (2022) [65]	Case study	Identifies disaster-related tweets to propose a structured, multi-layered approach for extracting actionable and relevant SA specifically tailored to the diverse informational needs of disaster responders and humanitarian organizations
Lowe et al (2016) [66]	Emergency medicine training	Explores the concept of SA within emergency medicine, focusing on developing a shared mental model among team members to enhance training and assessment methods
Khairilmizal et al (2023) [67]	Conceptual	Emphasizes effective disaster management, particularly in a sustainable context, hinges on robust strategies in 3 interconnected areas: Communication, Awareness, and Resource Management
De Monnin (2024) [68]	Conceptual and educational	Directly addresses the question of whether SA is an innate ability or a teachable skill. Likely explores pedagogical approaches and strategies for training SA
Westman et al (2024) [69]	Review of existing literature	Scopes the literature, and identifies SA as highly likely to be a prominent nontechnical skill
American College of Emergency Physicians (2014) [70]	Educational and informative	Outlines key competencies and identifies that SA is a fundamental competency in disaster response and likely included
Sapateiro and Antunes (2009) [71]	Conceptual	Proposes an emergency response model aimed at improving and enhancing SA
Lateef F (2022) [72]	Emergency medicine training	Explores the interconnectedness of SA, clinical reasoning, and clinical judgment in emergency medicine, advocating for cross-training using case-based discussions
Levin and Sauer (2012) [73]	Emergency medicine	Discusses the concept and importance of SA specifically within the context of emergency medicine. Likely covers its application and challenges in this setting
Bouzidi et al (2022) [74]	Case study (COVID-19)	Uses the COVID-19 pandemic as a case study to demonstrate how warning systems, SA, assessment methods, and education can be enhanced for emergency management
Lambert et al (2021) [75]	Technological development (AI)	Focuses on the application of AI to enhance SA during disaster response, discussing potential benefits and challenges
The Hospital Health System Association of Pennsylvania (2025) [76]	Policy and practical guidelines	Discusses the continued relevance of SA for hospitals and health care systems
Government Technology Handbook (2022) [77]	Policy and conceptual setting	Advocates for a new mindset in emergency response, integrating technology, data, and collaborative approaches to improve overall effectiveness, including SA

XSL•FO
RenderX

Study details and reference number	Study setting	Key points
Kuziemsky et al (2012) [78]	Information system design	Proposes an "upstream-downstream" approach for designing disaster management information systems, likely emphasizing the flow of information to support SA and decision-making
Laurila-Pant et al (2012) [79]	Simulation research methodology	Proposes a structured method for evaluating how well different agencies or teams develop a common understanding of a disaster situation and how this impacts their joint decision-making during simulated disaster exercises
Cooper et al (2013) [80]	Review of existing literature	Systematically reviews the various tools and methods used to measure SA in emergency settings, as well as the outcomes associated with SA
Williams et al. (2022) [81]	Paramedicine education	A pilot study investigating the impact of immersive simulation on paramedicine students' SA, using a mixed methods approach to assess its effectiveness
Gu et al (2018) [82]	Model development (Sensors)	Proposes a model for health and safety, SA, and emergency management that leverages multisensor signal fusion to gather comprehensive data and enhance understanding

[a]SA: situational awareness.

[b]DMS: disaster mindset.

[c]RFID: Radio Frequency Identification.

[d]UAV: unmanned aerial vehicle.

[e]SDN: software-defined networking.

[f]AI: artificial intelligence.

[g]SAR: Search and Rescue.

[h]EHR: electronic health record.

The Summary of the Results

Across the compiled documents, a consistent emphasis emerged on the pivotal role of SA in enhancing emergency response across various domains, including medical emergencies, disaster management, and multiagency operations. The number of documents discussing DMS was limited. These studies [24,25,36-82] indicate that access to real-time, accurate, and integrated data is fundamental (Table 1). Technologies such as unmanned aerial vehicles (UAVs), AI, and the Internet of Things (IoT) are transforming disaster response. However, the success of any system hinges on its usability and relevance to end users. Optimizing information flow, leveraging AI's growing role, and prioritizing real-time processing are vital for enhancing SA and improving outcomes in critical situations [24,25,36-82].

While technology is crucial in providing access to real-time information, it must be combined with purposeful training, clear communication, and an understanding of human cognitive factors. Furthermore, the transition from simply having information to making effective decisions is a critical area of focus. By addressing the challenges and leveraging the opportunities presented by technology and research, the ability to respond can significantly improve and mitigate the impact of disasters. Furthermore, the studies underscore the importance of integrating SA with other crucial skills such as clinical reasoning, communication, and teamwork, particularly in high-stress environments. Developing a DMS and shared mental models among responders is essential for effective coordination and adaptability. The research also reveals the limitations of traditional SA measurement methods and advocates for adopting comprehensive, mixed methods approaches to accurately assess and improve SA. Ultimately, the cumulative data advocate for a holistic, technology-enhanced, and competency-based strategy to cultivate and maintain SA, leading to improved emergency response outcomes and enhanced safety for both responders and affected populations [24,25,36-41,41-82].

This necessitates developing and implementing robust SA frameworks, tools, and training programs to ensure prompt and accurate information dissemination, thereby facilitating informed decision-making and efficient resource allocation. Technological advancements, such as AI-driven systems, multisensor signal fusion, and immersive simulations, are identified as critical components in improving SA by providing real-time data and enhancing responder capabilities. Only a few studies in the included publications discussed a DMS [25,77], explicitly addressing the concept of a "disaster readiness mindset," which inherently involves proactivity. The COVID-19 pandemic was used as a case study to emphasize the importance of being prepared and proactive in the face of potential disasters.

XSL•FO
RenderX

Discussion

Principal Findings

While the scientific studies consistently demonstrate that enhancing SA is paramount for effective disaster response and emergency management, the online reviews of several ongoing programs do not provide enough evidence on the extent to which these programs include the concepts of SA and a DMS. Most programs, however, touch both concepts indirectly by considering psychological, communicational, and decision-making aspects of disaster management. Table 2 provides information about a few current programs that seem to include SA and a DMS in their curricula [26-32].

XSL•FO
RenderX

Table . A short description of some disaster medicine courses that likely include situational awareness and a disaster mindset in their programs.

No. (Country)	Title	Description
1 (Spain, Sweden, Cyprus)	Erasmus Mundus Master Program - Public Health in Disasters (University of Oviedo, Karolinska Institutet, University of Nicosia).	A 2-year master's program, including "Global Health and Disasters" and "Public Health Response in Health Crisis and Disasters," which likely touches upon the psychological and contextual aspects of disaster response [26].
2 (Italy)	Training Disaster Medicine Trainers (TdmT) - IFMSA and CRIMEDIM (University of Piemonte Orientale).	Offers medical students the knowledge and skills in disaster medicine and global health, including topics such as complex humanitarian emergencies and disaster risk management. While the description doesn't explicitly mention "disaster mindset," the focus on humanitarian emergencies suggests the inclusion of psychological preparedness and situational awareness [27].
3 (Denmark)	Master of Disaster Management - University of Copenhagen.	Includes courses such as "Preparedness and Response to Humanitarian Crises" and "Disaster Risk Management: From Theory to Practice," likely involving elements of understanding the disaster environment and its psychological challenges [28].
4 (U.S.A.)	Disaster Fellowship - Johns Hopkins Department of Emergency Medicine.	This fellowship includes public health training and disaster field work, implying the development of skills to understand and navigate complex disaster scenarios, which include situational awareness and mindset [29].
5 (U.S.A.)	Center for Disaster Medicine - New York Medical College.	This center offers operational and tactical medicine training, including elements of decision-making under stress and maintaining situational awareness in chaotic environments [30].
6 (Sweden)	PRAD-MED – Preparedness and Disaster Medicine (Sahlgrenska University Hospital and Gothenburg University)	This 6-week-long course, spread over a year, includes global, national, and regional aspects of disaster and public health emergencies' preparedness and management. It has a mix of international and national experts and particularly discusses the transition of care from a multicasualty perspective to mass casualty and disaster casualty scenarios. It is a hybrid course using tabletop exercises and modular simulation training [31].
7 (Italy, Belgium)	The European Master in Disaster Medicine (EMDM)	This is a 1-year, blended Advanced Master of Science program for health professionals. It combines a self-directed online course with a 2-week intensive residential session featuring debates and complex simulations (including VR)[a], culminating in a final online exam and a research thesis. The EMDM[b] aims to equip participants with comprehensive knowledge, practical skills, and a systemic approach to health disaster management, fostering competent and adaptable professionals for emergency and humanitarian health response through its competency-based, multidisciplinary, and internationally focused curriculum [32].

[a]VR: virtual reality.

[b]EMDM: European Master in Disaster Medicine.

This study indicates that effective disaster response hinges on real-time, accurate, and integrated data. While technology revolutionizes information flow, boosting SA and decision-making, it needs purposeful training, clear communication, and an understanding of human cognitive factors to translate data into action. Cultivating a DMS and shared mental models is crucial, as well as integration of SA with skills such as clinical reasoning and teamwork. Many programs indirectly touch on these concepts, but more explicit inclusion and evidence of integration are needed. Ultimately, a holistic, technology-driven, and competency-based approach is vital for enhanced emergency response and safety.

XSL•FO
RenderX

The Evolving Landscape of SA in Disaster Management

SA, the ability to perceive, comprehend, and project the current state of a dynamic environment, has emerged as a critical factor across diverse domains because it enables individuals and teams to make informed decisions, anticipate potential problems, and respond effectively to dynamic situations, particularly in disaster response, medical emergencies, and biosurveillance. The papers reviewed in this study consistently underscore that mere data availability is insufficient; it must be transformed into actionable information, supporting informed decision-making and effective action [36-38,41,42].

Technology plays a pivotal role in enhancing SA. UAVs provide rapid visual information, offering comprehensive overviews and improved hazard identification. Radio Frequency Identification (RFID) enables real-time tracking of personnel and assets, enhancing accountability and safety. Effective information visualization tools are crucial for presenting complex data in an understandable format, especially during time-sensitive situations. Satellite imagery offers real-time wide-area information, while ubiquitous video, through technologies such as "RealityFlythrough," allows responders to virtually navigate and assess disaster zones [39,40,44-52].

However, achieving optimal SA is not without its challenges. Information overload, particularly from social media, can overwhelm decision-makers. Ensuring data veracity, especially from diverse sources, remains a significant hurdle. The transition from awareness to action requires decision-makers to effectively translate information into timely and appropriate responses. Individual cognitive factors, such as stress and fatigue, can also impact SA [53].

Endsley 3-level model [36] (perception, comprehension, and projection) serves as a valuable framework for understanding and improving SA. Studies consistently use this model to analyze how individuals process information in complex environments. Furthermore, simulation exercises and training programs are essential for developing the necessary skills and experience to maintain SA, particularly when implementing new technologies [83].

Recent advancements in data-driven approaches and AI have further revolutionized SA. End-user centricity is paramount, ensuring systems are tailored to the specific needs and preferences of those using them. Data integration from diverse sources creates comprehensive situational understanding, especially in complex environments such as air cargo operations. AI's transformative potential lies in its ability to process vast datasets, identify patterns, and provide predictive insights, enhancing early warning systems and resource allocation [44,47-50,54,55].

Human Sensor Networks highlight the value of citizen participation in gathering real-time, localized data, complementing traditional sensor networks. Technology integration, encompassing UAVs, GIS, remote sensing, social media, IoT, and AI, is essential for a unified situational view. Optimizing information flow within response networks, enhancing end-user centricity, and leveraging UAVs for scene assessment are critical for effective disaster management. Scaling 911 texting, using social media for early detection, and integrating deep learning with unmanned ground vehicles further enhance response capabilities. China's emergency platforms exemplify the integration of advanced technologies for comprehensive SA [51,52,56-58].

Deep learning for UAVs, visual cloud computing with software-defined networking, and ultralightweight deep learning models such as TinyEmergencyNet improve real-time visual information processing. Chebyshev Transform-based trajectory prediction enhances accuracy in apps such as autonomous vehicles [48,49].

Cultivating a Resilient DMS: Synthesis of Key Attributes

The concept of a "disaster mindset" emerges as a crucial element in navigating the complexities of modern crises [25]. It transcends mere reaction, advocating for a proactive, adaptable, and informed approach to preparedness and resilience. Analyzing the relevant literature reveals a set of key attributes that define this critical mindset.

A fundamental aspect of a DMS is the ability to anticipate potential threats and vulnerabilities. A DMS shifts the focus from reactive responses to proactive measures and involves moving beyond reactive measures and implementing preventative strategies. Performance-based infrastructure resilience studies exemplify this, showcasing the importance of evaluating resilience-enhancing options before a disaster strikes [84].

Disasters often present diverse challenges, requiring flexible and responsive strategies. Cross-cultural mood perception research underscores the need to recognize and adapt to varying cultural responses during crises, demonstrating that empathy and understanding are integral to effective support [1,85]. A DMS is characterized by a commitment to continuous learning from past events, both successes and failures. The analysis of the Great East-Japan Earthquake emphasizes the importance of incorporating historical lessons into current preparedness strategies, ensuring that past mistakes are not repeated [86].

In an increasingly interconnected and digital world, vigilance against evolving threats is essential. The study on cybersecurity in the SADC region highlights the need to adapt to new technologies and address emerging vulnerabilities, particularly in the realm of cyberattacks. A core component of a DMS is the commitment to building robust systems and communities that can withstand shocks and recover quickly. This includes the ability to construct resilient infrastructure and foster community resilience. A DMS is grounded in data and facts. It is crucial to have the ability to gather, analyze, and use information to make the best possible decisions. This ensures that responses are based on accurate and timely information [87-89].

Integration of Technology and Mindset

The effective integration of technology into disaster management is contingent upon a robust DMS. While technology provides powerful tools for early warning, SA, and response, it is the

XSL•FO
RenderX

proactive, adaptable, and informed mindset that ensures these tools are used effectively. For example, AI can enhance predictive capabilities, but a DMS ensures that these predictions are used to inform proactive measures and build resilience. The DMS is not static; it must evolve in response to changing threats and technological advancements. The COVID-19 pandemic significantly increased cybersecurity awareness, highlighting the necessity for a flexible and responsive approach [25,88]. Similarly, the increase in climate-related disasters underscores the necessity for an evolving mindset that can address new and emerging challenges. Cultivating a resilient DMS is crucial for navigating the complexities of modern crises. It involves a proactive, adaptable, and informed approach to preparedness and resilience [89,90]. By embracing these key attributes, individuals, organizations, and communities can better anticipate, respond to, and recover from disasters, ultimately building a safer and more resilient future.

Using Technology to Fill in the Gaps in Disaster Simulation Exercises

Disaster exercises primarily focus on the core pillars of disaster medicine, including planning, response, safety, triage, clinical competence, psychological first aid, interdisciplinary collaboration, and continuous improvement. Nevertheless, the main challenge is to capture the dynamic of an incident that may transition from a multiple casualty to a mass casualty and further to a disaster casualty event. As emergencies escalate, this transition unfolds like increasingly challenging chapters in crisis response [19,90,91].

In a multiple casualty incident, numerous injured individuals are involved. Local emergency services, including paramedics, firefighters, and police, swiftly arrive to triage the injured and prioritize those needing immediate medical attention. The local hospital prepares to receive patients, and the situation remains manageable within the city's emergency resources [92,93]. In a mass casualty incident, the number of casualties quickly overwhelms local emergency services. Regional and national emergency response plans are activated, bringing additional medical teams, equipment, and supplies from neighboring areas. Coordination becomes complex as multiple agencies work together, focusing on systematic triage to use resources efficiently and save lives [94]. Finally, in a disaster casualty situation, a catastrophic event devastates an entire region, affecting infrastructure, homes, and critical services. The number of casualties is overwhelming, exhausting local and regional resources. A comprehensive, multisectoral approach is required, involving national disaster response plans and international aid. Coordination expands to include government agencies, nongovernmental organizations (NGOs), international relief organizations, and community leaders, addressing immediate medical care and long-term recovery and rebuilding efforts [94,95].

Throughout these transitions, the DMS and SA evolve. In a multiple casualty incident, the focus is on rapid response and localized resource management. As the situation escalates to a mass casualty incident, the response becomes scalable, requiring broader coordination and systematic triage. Finally, in a disaster casualty situation, the response is comprehensive, involving strategic coordination across multiple sectors and long-term planning. While current simulation exercises in disaster medicine are invaluable for training and preparedness, they have several gaps, which influence the outcomes of simulation exercises and the development of SA and DMS in individuals and collectively [38,91-95].

These gaps must be addressed to enhance their effectiveness. Table 3 shows the gaps in disaster medicine simulation and the ideal technological solutions. These solutions can help close the gaps in current disaster medicine simulation exercises, enhancing the training and preparedness of health care professionals [95-98].

XSL•FO
RenderX

Table . Outlines the gaps in disaster medicine simulation and the corresponding technological solutions.

Gaps	Description	Technological solutions
Lack of uniform standards	There is a need for standardized procedures and protocols across different simulation exercises to ensure consistent training outcomes.	Advanced simulation software to standardize procedures and protocols
Insufficient rigorous evaluation	Many exercises lack rigorous evaluation methods to comprehensively assess their effectiveness, making it difficult to identify areas for improvement.	AI[a] for automated evaluation and detailed feedback
Limited feedback mechanisms	Immediate and constructive feedback is crucial for learning, but some exercises do not provide adequate feedback to participants.	AI-driven systems for real-time, constructive feedback
Limited realism	Some simulations may not fully capture the complexity and unpredictability of actual disasters, limiting participants' preparedness for real-life situations.	VR[b] and AR[c] for highly realistic and immersive disaster scenarios
Scenario diversity	There is often a lack of diverse scenarios covering a wide range of potential disasters, including rare but high-impact events.	VR and AR to simulate a wide range of disaster scenarios
Gaps in interagency coordination	Effective disaster response requires seamless coordination between various agencies and disciplines, which some exercises do not adequately address.	Telemedicine platforms for seamless communication and coordination
Limited resources	High-quality simulation exercises can be resource-intensive, requiring significant time, personnel, and financial investment, limiting the frequency and scope of training.	mHealth apps for efficient resource management and cost-effective training
Underuse of advanced technologies	Technologies such as VR and AR have the potential to enhance simulation training but are not yet widely integrated into disaster medicine education.	Blended learning platforms, combining online modules with hands-on simulations

[a]AI: artificial intelligence.

[b]VR: virtual reality.

[c]AR: augmented reality.

Fostering SA and Developing a DMS

While UAVs, GIS, AI, IoT, and social media monitoring offer significant capabilities, the optimal technology for enhancing SA in disaster management is context dependent. A well-integrated system combining these tools, with a strong emphasis on data management and user-centered design, is likely to be most effective. Furthermore, technologies such as VR and AR simulations and online educational platforms can cultivate a proactive DMS through realistic training and widespread dissemination of SA. Data visualization promotes data-driven preparedness, while reliable communication and alert systems ensure timely information sharing. Social media and crowdsourcing can also foster community resilience and shared responsibility. However, the successful adoption and impact of these technologies, crucial for building a robust DMS, hinge on accessibility, information accuracy, and adequate user training [25,45,46,51,53,96-101].

Indeed, research in information systems underscores the critical role of user adoption, often explained by the Technology Acceptance Model (TAM), which posits that perceived usefulness and ease of use are key determinants of whether technologies are embraced or discarded [102].

Finally, effective education for SA and DMS must adapt to diverse cultural, social, and economic contexts. This means respecting local beliefs, integrating traditional knowledge, and using community leaders as trusted communicators, often through visual or oral methods. Socially, programs should be co-created with communities, leverage existing networks, and explicitly address vulnerabilities of marginalized groups to build trust and inclusivity [103]. Furthermore, the type of disaster significantly impacts educational needs. For slow-onset events such as drought or sea-level rise, initiatives focus on long-term preparedness, sustainable practices, and climate adaptation. For sudden-onset events such as earthquakes or tsunamis, the emphasis shifts to immediate evacuation protocols, early warning system comprehension, and rapid decision-making skills. Education for man-made disasters (eg, industrial accidents and terrorism) would include understanding specific hazards, lockdown procedures, or emergency contact protocols, emphasizing different aspects of SA and mindset than natural hazards [104,105]. Table 3 shows the gaps in disaster education and how technology can be used as a solution.

Limitations

Unlike systematic reviews, this study did not assess the quality or risk of bias of the included studies. This means that some

XSL•FO
RenderX

findings, while providing an overview, might be influenced by methodologically weaker research.

As a scoping review, the aim of this study was to map the breadth of available literature rather than delve into the in-depth nuances of individual studies or provide definitive answers on intervention effectiveness. While we used a systematic process, and 3 researchers mitigated potential subjectivity during study selection and data extraction, the broad inclusion of diverse methodologies, populations, and outcomes naturally presents challenges in synthesizing findings into a singular quantitative or qualitative conclusion.

A significant limitation arises from the sole inclusion of studies in English. Due to this language restriction, valuable insights from non-English literature might have been missed. With the increasing sophistication and accessibility of AI translation tools, relying solely on English-language publications could be a substantial limitation. While AI translation is improving, it still grapples with accurately capturing idiomatic expressions, cultural nuances, and complex contextual meanings, particularly in specialized fields such as disaster medicine. Therefore, the absence of non-English studies, even with the potential of AI translation, could introduce a language and cultural bias, potentially limiting the comprehensiveness and global generalizability of this review's findings.

Future Directions

The future of leveraging technology for SA and DMS is heading toward deep integration, automation, and intelligent systems, shifting us to a more proactive and adaptive approach in disaster management [105]. Key trends include hyperenhanced SA through ubiquitous IoT sensors and AI-powered data analysis, allowing for real-time prediction and actionable insights. Immersive visualization via AR, VR, and MR will revolutionize how decision-makers perceive disaster environments, while digital twins will enable dynamic scenariosimulations and optimized resource allocation [13,96].

Crucially, technology will also foster a more proactive and adaptive DMS. This involves personalized preparedness through wearables and AI, alongside immersive training in VR and AR simulations that build psychological resilience. AI-augmented decision support will act as an intelligent copilot for emergency managers, ensuring seamless human-AI collaboration. While challenges such as interoperability, data security, and ethical considerations remain, the overarching goal is to create a highly interconnected, intelligent, and immersive technological ecosystem that significantly enhances our collective ability to anticipate, respond to, and recover from disasters [13,96].

Conclusions

In conclusion, the role of technology in enhancing situational awareness and fostering a robust DMS is unequivocal yet necessitates careful consideration. While tools such as UAVs, AI, and the IoT offer unprecedented data and analytical capabilities, their effectiveness is contingent upon user-centered design, data integrity, and seamless integration. Importantly, cultivating a proactive DMS extends beyond the mere deployment of technology. VR and AR simulations, educational platforms, and communication systems are instrumental in developing preparedness, resilience, and informed decision-making. Ultimately, the successful application of technology in disaster management requires a holistic approach that integrates advanced tools with human-centered design, effective information management, and a commitment to continuous learning. Moving beyond simply equipping individuals and teams with technical solutions, the true significance lies in their ability to use these tools effectively, ensuring they are also mentally prepared to navigate the complexities of crises and can seamlessly integrate technology into their response strategies.

Data Availability

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

Authors' Contributions

AKM conceptualized the study. AKM, MRJ, and LRN reviewed and selected the included studies. AKM drafted the first version. All authors (AKM, MRJ, LRN, EC, LB, LS, and JLS) reviewed, edited, and approved the final version to be submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search results from scientific databases.
[PDF File, 168 KB - ojphi_v17i1e75404_app1.pdf]

Multimedia Appendix 2
Search results from Google Scholar search.
[PDF File, 382 KB - ojphi_v17i1e75404_app2.pdf]

Checklist 1
PRISMA-ScR checklist.

XSL•FO
RenderX

[PDF File, 748 KB - ojphi_v17i1e75404_app3.pdf]

References

1. Khorram-Manesh A. Flexible surge capacity - public health, public education, and disaster management. Health Promot Perspect 2020;10(3):175-179. [doi: 10.34172/hpp.2020.30] [Medline: 32802753]
2. Wu TC, Ho CTB. A scoping review of metaverse in emergency medicine. Australas Emerg Care 2023 Mar;26(1):75-83. [doi: 10.1016/j.auec.2022.08.002] [Medline: 35953392]
3. Shaw R. Thirty years of science, technology, and academia in disaster risk reduction and emerging responsibilities. Int J Disaster Risk Sci 2020 Aug;11(4):414-425. [doi: 10.1007/s13753-020-00264-z]
4. Ali K, Nguyen HX, Shah P. Disaster management communication networks: challenges and architecture design. Presented at: 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops); Mar 23-27, 2015; St Louis, MO p. 537-542. [doi: 10.1109/PERCOMW.2015.7134094]
5. Abbas SH, Srivastava RK, Tiwari RP, Bala Ramudu P. GIS‐based disaster management: a case study for Allahabad Sadar sub‐district (India). Manag Environ Qual Int J 2009;20(1):33-51. [doi: 10.1108/14777830910922433]
6. Voigt S, Kemper T, Riedlinger T, Kiefl R, Scholte K, Mehl H. Satellite image analysis for disaster and crisis-management support. IEEE Trans Geosci Remote Sensing 2007;45(6):1520-1528. [doi: 10.1109/TGRS.2007.895830]
7. Xiao Y, Huang Q, Wu K. Understanding social media data for disaster management. Nat Hazards 2015 Dec;79(3):1663-1679. [doi: 10.1007/s11069-015-1918-0]
8. Arabboev M, Begmatov S, Nosirov K, Chedjou JC, Kyamakya K. Development of a wearable device for monitoring and predicting human health in emergencies. Tech Sci Innov 2022(1):67-73 [FREE Full text]
9. Khanal S, Medasetti US, Mashal M, Savage B, Khadka R. Virtual and augmented reality in the disaster management technology: a literature review of the past 11 years. Front Virtual Real 2022;3:843195. [doi: 10.3389/frvir.2022.843195]
10. Linardos V, Drakaki M, Tzionas P, Karnavas YL. Machine learning in disaster management: recent developments in methods and applications. MAKE 2022;4(2):446-473. [doi: 10.3390/make4020020]
11. Anthony Jnr B. Implications of telehealth and digital care solutions during COVID-19 pandemic: a qualitative literature review. Inform Health Soc Care 2021 Mar 2;46(1):68-83. [doi: 10.1080/17538157.2020.1839467] [Medline: 33251894]
12. Phattharapornjaroen P, Carlström E, Sivarak O, et al. Community-based response to the COVID-19 pandemic: case study of a home isolation centre using flexible surge capacity. Public Health 2022 Oct;211:29-36. [doi: 10.1016/j.puhe.2022.06.025] [Medline: 35994836]
13. Krichen M, Abdalzaher MS, Elwekeil M, Fouda MM. Managing natural disasters: an analysis of technological advancements, opportunities, and challenges. Internet of Things and Cyber-Physical Systems 2024;4:99-109. [doi: 10.1016/j.iotcps.2023.09.002]
14. Khorram-Manesh A, Lupesco O, Friedl T, et al. Education in disaster management: what do we offer and what do we need? Proposing a new global program. Disaster Med Public Health Prep 2016 Dec;10(6):854-873. [doi: 10.1017/dmp.2016.88] [Medline: 27435533]
15. Gordon M, Daniel M, Ajiboye A, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach 2024 Apr;46(4):446-470. [doi: 10.1080/0142159X.2024.2314198] [Medline: 38423127]
16. Wexler BJ, Schultz C, Biddinger PD, et al. The 2023 model core content of disaster medicine. Prehosp Disaster Med 2023 Dec;38(6):699-706. [doi: 10.1017/S1049023X23006556] [Medline: 37869875]
17. Tin D, Hertelendy AJ, Ciottone GR. Disaster medicine training: the case for virtual reality. Am J Emerg Med 2021 Oct;48:370-371. [doi: 10.1016/j.ajem.2021.01.085] [Medline: 33581938]
18. Khorram-Manesh A, Berlin J, Carlström E. Two validated ways of improving the ability of decision-making in emergencies; results from a literature review. Bull Emerg Trauma 2016 Oct;4(4):186-196. [Medline: 27878123]
19. Sultan MAS, Carlström E, Sørensen JL, Alruwaili AS, Khorram‐Manesh A. Incorporating simulation exercises using collaborative tools into disaster and emergency medicine curriculum—a pilot survey among Saudi Arabian professionals. Contingencies & Crisis Mgmt 2023 Dec;31(4):905-912. [doi: 10.1111/1468-5973.12491]
20. Dillon PM, Noble KA, Kaplan L. Simulation as a means to foster collaborative interdisciplinary education. Nurs Educ Perspect 2009;30(2):87-90. [Medline: 19476071]
21. Shevellar L, Riggs R. Understanding resistance to emergency and disaster messaging. Aust J Emerg Manag 2015;30(3):31-35 [FREE Full text]
22. Rabi M. The rise of global health emergency governance. Minerva (Arlingt Va) 2025 Sep;63(3):471-497. [doi: 10.1007/s11024-024-09555-x]
23. Bertuzzi N. Rethinking crisis management: technocracy, globalism, and the rise of emergenceocracy. SSRN. Preprint posted online on Feb 7, 2025. [doi: 10.2139/ssrn.5117957]
24. Mohsin B, Steinhäusler F, Madl P, Kiefel M. An innovative system to enhance situational awareness in disaster response. J Homeland Secur Emerg Manag 2016 Sep 1;13(3):301-327. [doi: 10.1515/jhsem-2015-0079]
25. Su Z, McDonnell D, Ahmad J. The need for a disaster readiness mindset: a key lesson from the coronavirus disease 2019 (COVID-19) pandemic. Infect Control Hosp Epidemiol 2022 Apr;43(4):538-539. [doi: 10.1017/ice.2021.26] [Medline: 33487209]

XSL•FO
RenderX

26. Erasmus Mundus Master Programme - Public Health in Disasters (University of Oviedo, Karolinska Institutet, University of Nicosia). URL: https://masterpublichealthdisasters.eu/ [accessed 2025-04-02]
27. Call for the participants for the training disaster medicine trainers 2025. IFMSA. URL: https://ifmsa.org/call-for-participants-for-the-training-disaster-medicine-trainers-2025/ [accessed 2025-04-02]
28. Master of Disaster Management. University of Copenhagen. URL: https://continuing-education.ku.dk/master-of-disaster-management/ [accessed 2025-04-02]
29. Disaster Fellowship - Johns Hopkins Department of Emergency Medicine. Johns Hopkins Medicine. URL: https://www.hopkinsmedicine.org/emergency-medicine/fellowship-programs/disaster [accessed 2025-04-02]
30. Center for Disaster Medicine. New York Medical College. URL: https://www.nymc.edu/research/centers--institutes/center-for-disaster-medicine/ [accessed 2025-04-02]
31. Programs in emergency preparedness and disaster medicine. Sahlgrenska University Hospital. URL: https://www.sahlgrenska.se/jobb-och-framtid/student-och-utbildningstjanst/st/Program-inom-beredskap-och-katastrofmedicin/ [accessed 2025-04-02]
32. European Master in Disaster Medicine. UPO - CRIMEDIM. URL: https://crimedim.uniupo.it/european-master-in-disaster-medicine [accessed 2025-07-11]
33. Gusenbauer M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. Scientometrics 2019 Jan;118(1):177-214. [doi: 10.1007/s11192-018-2958-5]
34. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Dec;18(1):1-7. [doi: 10.1186/s12874-018-0611-x]
35. Sabharwal M, Levine H, D'Agostino M. A conceptual content analysis of 75 years of diversity research in public administration. Rev Pub Pers Adm 2018 Jun;38(2):248-267. [doi: 10.1177/0734371X16671368]
36. Naor M, Laor E. Disaster recovery after Hurricane Maria in Puerto Rico: assessment using Endsley's three-level model of situational awareness. J Bus Contin Emer Plan 2020;13(3):278. [doi: 10.69554/BVIX6612]
37. Kettelhut VV, Vanschooneveld TC, McClay JC, Mercer DF, Fruhling A, Meza JL. Empirical study on the impact of a tactical biosurveillance information visualization on users' situational awareness. Mil Med 2017 Mar;182(S1):322-329. [doi: 10.7205/MILMED-D-16-00143]
38. Glick JA, Barbara JA. Moving from situational awareness to decisions during disaster response: transition to decision making. J Emerg Manag 2013;11(6):423-432. [doi: 10.5055/jem.2013.0155] [Medline: 24623111]
39. Jokela J, Rådestad M, Gryth D, et al. Increased situation awareness in major incidents-radio frequency identification (RFID) technique: a promising tool. Prehosp Disaster Med 2012 Feb;27(1):81-87. [doi: 10.1017/S1049023X12000295] [Medline: 22575228]
40. McCurdy NJ, Griswold WG, Lenert LA. RealityFlythrough: enhancing situational awareness for medical response to disasters using ubiquitous video. AMIA Annu Symp Proc 2005;2005:510-514. [Medline: 16779092]
41. Kedia T, Ratcliff J, O'Connor M, et al. Technologies enabling situational awareness during disaster response: a systematic review. Disaster Med Public Health Prep 2022 Feb;16(1):341-359. [doi: 10.1017/dmp.2020.196] [Medline: 32829725]
42. Saadou A, Chenji H. Optimizing situational awareness in disaster response networks. IEEE Access 2018;6:24625-24638. [doi: 10.1109/ACCESS.2018.2831448]
43. Yuan W, Guan D, Huh EN, Lee S. Harness human sensor networks for situational awareness in disaster reliefs: a survey. IETE Tech Rev 2013;30(3):240. [doi: 10.4103/0256-4602.113522]
44. Karavarsamis S, Gkika I, Gkitsas V, Konstantoudakis K, Zarpalas D. A survey of deep learning-based image restoration methods for enhancing situational awareness at disaster sites: the cases of rain, snow and haze. Sensors (Basel) 2022 Jun 22;22(13):4707. [doi: 10.3390/s22134707] [Medline: 35808203]
45. Jain T, Sibley A, Stryhn H, Lund A, Hubloue I. Comparison of unmanned aerial vehicle technology versus standard practice of scene assessment by paramedic students of a mass-gathering event. Prehosp Disaster Med 2021 Dec;36(6):756-761. [doi: 10.1017/S1049023X2100114X] [Medline: 34702422]
46. Jain T, Sibley A, Stryhn H, Hubloue I. Comparison of unmanned aerial vehicle technology versus standard practice in identification of hazards at a mass casualty incident scenario by primary care paramedic students. Disaster Med Public Health Prep 2018 Oct;12(5):631-634. [doi: 10.1017/dmp.2017.129] [Medline: 29382398]
47. Lee MFR, Chen YC, Tsai CY. Deep learning-based human body posture recognition and tracking for unmanned aerial vehicles. Processes (Basel) 2022;10(11):2295. [doi: 10.3390/pr10112295]
48. Mogaka OM, Zewail R, Inoue K, Sayed MS. TinyEmergencyNet: a hardware-friendly ultra-lightweight deep learning model for aerial scene image classification. J Real-Time Image Proc 2024 Apr;21(2):51. [doi: 10.1007/s11554-024-01430-y]
49. Spandonidis C, Giannopoulos F, Petsa A, Eleftheridis P, Sedikos E. A data-driven situational awareness system for enhanced air cargo operations emergency control. Smart Cities 2021;4(3):1087-1103. [doi: 10.3390/smartcities4030057]
50. Yuan H, Huang L, Chen T, et al. Key technologies of the emergency platform in China. J Saf Sci Resil 2022 Dec;3(4):404-415. [doi: 10.1016/j.jnlssr.2022.10.001]

XSL•FO
RenderX

51. Tapia AH, Giacobe NA, Soule PJ, LaLone NJ. Scaling 911 texting for large-scale disasters: developing practical technical innovations for emergency management at public universities. Int J Public Adm Digital Age 2016;3(3):73-85. [doi: 10.4018/IJPADA.2016070105]
52. Gargees R, Morago B, Pelapur R, et al. Incident-supporting visual cloud computing utilizing software-defined networking. IEEE Trans Circuits Syst Video Technol 2016;27(1):182-197. [doi: 10.1109/TCSVT.2016.2564898]
53. Chan JL, Purohit H. Challenges to transforming unconventional social media data into actionable knowledge for public health systems during disasters. Disaster Med Public Health Prep 2020 Jun;14(3):352-359. [doi: 10.1017/dmp.2019.92] [Medline: 31610817]
54. Abid SK, Sulaiman N, Chan SW, et al. Toward an integrated disaster management approach: how artificial intelligence can boost disaster management. Sustainability 2021;13(22):12560. [doi: 10.3390/su132212560]
55. Rasouli Panah H, Madanian S, Yu J. Disaster health care and resiliency: a systematic review of the application of social network data analytics. Disaster Med Public Health Prep 2024;18:e334. [doi: 10.1017/dmp.2024.294]
56. Pekar V, Binner J, Najafi H, Hale C, Schmidt V. Early detection of heterogeneous disaster events using social media. Asso for Info Science & Tech 2020 Jan;71(1):43-54. [doi: 10.1002/asi.24208]
57. Zafar MH, Moosavi SKR, Sanfilippo F. Enhancing unmanned ground vehicle performance in SAR operations: integrated gesture-control and deep learning framework for optimised victim detection. Front Robot AI 2024;11:1356345. [doi: 10.3389/frobt.2024.1356345] [Medline: 38957217]
58. Lenert LA, Kirsh D, Griswold WG, et al. Design and evaluation of a wireless electronic health records system for field care in mass casualty settings. J Am Med Inform Assoc 2011;18(6):842-852. [doi: 10.1136/amiajnl-2011-000229] [Medline: 21709162]
59. Gao L, Song C, Gao Z, Barabási AL, Bagrow JP, Wang D. Quantifying information flow during emergencies. Sci Rep 2014 Feb 6;4(1):3997. [doi: 10.1038/srep03997] [Medline: 24499738]
60. Gandhi S, Yeager J, Glaman R. Implementation and evaluation of a pandemic simulation exercise among undergraduate public health and nursing students: a mixed-methods study. Nurse Educ Today 2021 Mar;98:104654. [doi: 10.1016/j.nedt.2020.104654] [Medline: 33218905]
61. Stroud C, Altevogt BM, Nadig L, Hougan M. Value-based models for sustaining emergency preparedness capacity and capability in the United States. : Institute of Medicine (US) Forum on Medical and Public Health Preparedness for Catastrophic Events; 2010.
62. Iserson KV. Integrating disaster response tools for clinical leadership. West J Emerg Med 2024;26(1). [doi: 10.5811/WESTJEM.35390]
63. O'Brien A, Read GJM, Salmon PM. Situation awareness in multi-agency emergency response: models, methods and applications. Int J Disaster Risk Reduct 2020 Sep;48:101634. [doi: 10.1016/j.ijdrr.2020.101634]
64. Schwarz K, Aranda DA, Hartmann M. Towards automated situational awareness reporting for disaster management—a case study. Sustainability 2023;15(10):7968. [doi: 10.3390/su15107968]
65. Zhai W. A multi-level analytic framework for disaster situational awareness using Twitter data. Comput Urban Sci 2022;2(1):23. [doi: 10.1007/s43762-022-00052-z]
66. Lowe DJ, Ireland AJ, Ross A, Ker J. Exploring situational awareness in emergency medicine: developing a shared mental model to enhance training and assessment. Postgrad Med J 2016 Nov 1;92(1093):653-658. [doi: 10.1136/postgradmedj-2015-133772]
67. Khairilmizal S, Hussin MF, Alabdouli K. Criticality of sustainable disaster management through communication, awareness, and resource management. Journal of Southwest Jiaotong University 2023;58(2) [FREE Full text] [doi: 10.35741/issn.0258-2724.58.2.70]
68. Monnin C. Can situational awareness be taught? National Association of EMS Physicians. URL: https://naemsp.org/can-situational-awareness-be-taught/ [accessed 2025-04-02]
69. Westman A, Kurland L, Hugelius K. Non-technical skills needed by medical disaster responders– a scoping review. Scand J Trauma Resusc Emerg Med 2024;32(1):25. [doi: 10.1186/s13049-024-01197-y]
70. American College of Emergency Physicians. Development of competencies for disaster medicine fellowships. an information paper. : American College of Emergency Physicians; 2014 URL: https://tinyurl.com/3e2ukrrw [accessed 2025-04-02]
71. Sapateiro C, Antunes P. An emergency response model toward situational awareness improvement. Presented at: International conference on information systems for crisis response and management; May 10-13, 2009; Göteborg, Sweden.
72. Lateef F. Tying in situational awareness, clinical reasoning and clinical judgment through cross-training using case-based discussions in emergency medicine. Curr Res Emerg Med 2022;2(2):1-5. [doi: 10.54026/CREM/1024]
73. Levin S, Sauer L, Kelen G, et al. Situation awareness in emergency medicine. IIE Trans Healthc Syst Eng 2012 Apr;2(2):172-180. [doi: 10.1080/19488300.2012.684739]
74. Bouzidi Z, Amad M, Boudries A. Enhancing warning, situational awareness, assessment and education in managing emergency: case study of COVID-19. SN Comput Sci 2022;3(6):454. [doi: 10.1007/s42979-022-01351-2] [Medline: 36035507]
75. Lambert JD, Rose M, Ratcliff J, et al. AI-enabled situational awareness in disaster response. Defense Technical Information Center. 2021. URL: https://apps.dtic.mil/sti/citations/AD1143415 [accessed 2025-04-02]

XSL•FO
RenderX

76. Why COVID-19 situational awareness still matters as national emergency ends. The Hospital Healthsystem Association of Pennsylvania. 2024. URL: https://www.haponline.org/News/HAP-News-Articles/HAP-Blog/why-covid-19-situational-awareness-still [accessed 2025-04-02]
77. A Government Technology Handbook. Verizon Frontline. A new mindset for emergency response. how state and local governments can use technology to create more sustainable, reliable, and resilient operations. 2022 URL: https://media.erepublic.com/document/GT22_HANDBOOK_Verizon_EM_V.pdf [accessed 2025-04-02]
78. Kuziemsky CE, O'Sullivan TL, Corneil W. An upstream-downstream approach for disaster management information systems design. In: ISCRAM 2012 Conference Proceedings - 9th International Conference on Information Systems for Crisis Response and Management: Simon Fraser University; 2012:1-10.
79. Laurila-Pant M, Pihlajamäki M, Lanki A, Lehikoinen A. A protocol for analysing the role of shared situational awareness and decision-making in cooperative disaster simulations. Int J Disaster Risk Reduct 2023 Feb;86:103544. [doi: 10.1016/j.ijdrr.2023.103544]
80. Cooper S, Porter J, Peach L. Measuring situation awareness in emergency settings: a systematic review of tools and outcomes. Open Access Emerg Med 2014;6:1-7. [doi: 10.2147/OAEM.S53679] [Medline: 27147872]
81. Williams B, Beovich B, Olaussen A, Hunter J. Immersive imulation and paramedicine students' situational awareness: a mixed-methods pilot study. MRAJ 2022;10(9). [doi: 10.18103/mra.v10i9.3029]
82. Gu Q, Jiang S, Lian M, Lu C. Health and safety situation awareness model and emergency management based on multi-sensor signal fusion. IEEE Access 2018;7:958-968. [doi: 10.1109/ACCESS.2018.2886061]
83. Endsley MR. Situation awareness. In: Handbook of Human Factors and Ergonomics: John Wiley & Sons, Inc; 2021:434-455. [doi: 10.1002/9781119636113]
84. Norris FH, Stevens SP, Pfefferbaum B, Wyche KF, Pfefferbaum RL. Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. Am J Community Psychol 2008 Mar;41(1-2):127-150. [doi: 10.1007/s10464-007-9156-6] [Medline: 18157631]
85. Lucini B. Multicultural approaches to disaster and cultural resilience. How to consider them to improve disaster management and prevention: the Italian case of two earthquakes. Procedia Econ Financ 2014;18:151-156. [doi: 10.1016/S2212-5671(14)00925-3]
86. Takemoto S, Shibuya N, Sakoda K. Learning from mega-disasters: a decade of lessons from the great east japan earthquake. PreventionWeb. URL: https://www.preventionweb.net/news/learning-megadisasters-decade-lessons-great-east-japan-earthquake [accessed 2025-04-02]
87. Stewart J. Critical infrastructure systems are vulnerable to new kinds of cyberattacks. Georgia Tech Published February.: College of Engineering; 2024 Feb 2. URL: https://tinyurl.com/3r34m29a [accessed 2024-04-02]
88. Bagui L, Lusinga S, Pule N, et al. The impact of COVID‐19 on cybersecurity awareness‐raising and mindset in the southern African development community (SADC). Electron J Inf Syst Dev Ctries 2023;89(4):e12264. [doi: 10.1002/isd2.12264]
89. Merrick B. Changing mindset, perceptions, learning, and tradition: an "adaptive teaching framework" for teaching music online. Int J Innov Online Edu 2020;4(2). [doi: 10.1615/IntJInnovOnlineEdu.2020035150]
90. Goniewicz K, Burkle FM, Khorram-Manesh A. Transforming global public health: climate collaboration, political challenges, and systemic change. J Infect Public Health 2025 Jan;18(1):102615. [doi: 10.1016/j.jiph.2024.102615] [Medline: 39644717]
91. Kritikos N, Grant J. Simulation in healthcare education: what are the essential standards. World J Methodol 2013;3(1):10. [doi: 10.1002/9781118748039]
92. Sloan HM. Responding to a multiple-casualty incident: room for improvement. J Emerg Nurs 2011 Sep;37(5):484-486. [doi: 10.1016/j.jen.2010.07.013] [Medline: 21889656]
93. Busby S, Witucki-Brown J. Theory development for situational awareness in multi-casualty incidents. J Emerg Nurs 2011 Sep;37(5):444-452. [doi: 10.1016/j.jen.2010.07.023] [Medline: 21889652]
94. Lynn M, Gurr D, Memon A, Kaliff J. Management of conventional mass casualty incidents: ten commandments for hospital planning. J Burn Care Res 2006;27(5):649-658. [doi: 10.1097/01.BCR.0000238119.29269.2B] [Medline: 16998397]
95. Chauvin C, Clostermann JP, Hoc JM. Situation awareness and the decision-making process in a dynamic situation: avoiding collisions at sea. J Cogn Eng Decis Mak 2008 Mar;2(1):1-23. [doi: 10.1518/155534308X284345]
96. Khorram-Manesh A, Shin H, Carlström E. Impact of technology on disaster preparedness and healthcare management. In: Augusto JC, editor. Handbook on Smart Health Studies in Health Technology and Informatics Book Series: Sage/IOS Press; 2025.
97. Ardito L, Coccia M, Messeni Petruzzelli A. Technological exaptation and crisis management: evidence from COVID‐19 outbreaks. R & D Management 2021 Sep;51(4):381-392. [doi: 10.1111/radm.12455]
98. Zhu Y, Li N. Virtual and augmented reality technologies for emergency management in the built environments: a state-of-the-art review. J Saf Sci Resil 2021 Mar;2(1):1-10. [doi: 10.1016/j.jnlssr.2020.11.004] [Medline: 40477419]
99. Nazli N, Sipon S, Radzi HM. Analysis of training needs in disaster preparedness. Procedia Soc Behav Sci 2014 Aug;140:576-580. [doi: 10.1016/j.sbspro.2014.04.473]
100. Public health emergency preparedness program and guidance. Center of Disease Control. URL: https://www.cdc.gov/readiness/php/phep/index.html [accessed 2025-04-02]

XSL•FO
RenderX

101. Gunawan I, Afiantari F, Kusumaningrum DE, et al. Improving disaster response through disaster simulation. Int J Innov Creat Change 2019;5(4):640-653 [FREE Full text]
102. Drachsler H, Jansen J, Kirschner PA. Adoption of learning technologies in times of pandemic crisis. Computer Assisted Learning 2021 Dec;37(6):1509-1512. [doi: 10.1111/jcal.12626]
103. Chan G, Banire B, Ataguba G, Frempong G, Orji R. A panoramic view of socio-cultural sensitivity in digital technologies: a comprehensive review and future directions. Int J Hum–Comput Interact 2025 Feb 16;41(4):1917-1945. [doi: 10.1080/10447318.2024.2372135]
104. Guo L, Fang M, Liu L, Chong H, Zeng W, Hu X. The development of disaster preparedness education for public: a scoping review. BMC Public Health 2025 Feb 17;25(1):645. [doi: 10.1186/s12889-025-21664-0] [Medline: 39962440]
105. Khorram-Manesh A, Eskici GT, Gray L, et al. Enhancing global disaster preparedness: a scoping review of the current integration of situational awareness and disaster mindset in healthcare education. AIMS Public Health 2025;12(3):735-766.

Abbreviations

AI: artificial intelligence
AR: augmented reality
DMS: disaster mindset
GIS: geographic information systems
IoT: Internet of Things
NGO: nongovernmental organization
RFID: Radio Frequency Identification
SA: situational awareness
TAM: Technology Acceptance Model
UAV: unmanned aerial vehicles
VR: virtual reality

Edited by E Mensah; submitted 02.04.25; peer-reviewed by M Nakayama, Y Elsanousi; revised version received 08.09.25; accepted 22.09.25; published 10.10.25.

Please cite as:
Khorram-Manesh A, Johannessen MR, Nielsen LR, Carlström E, Berntzen L, Sandberg L, Sørensen JL
Cultivating Disaster Preparedness: Scoping Review of Technology's Contribution to Situational Awareness and Disaster Mindset in Disaster Medicine
Online J Public Health Inform 2025;17:e75404
URL: https://ojphi.jmir.org/2025/1/e75404
doi:10.2196/75404

© Amir Khorram-Manesh, Marius Rohde Johannessen, Laurits Rauer Nielsen, Eric Carlström, Lasse Berntzen, Lene Sandberg, Jarle Løwe Sørensen. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 10.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

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

XSL•FO
RenderX